10018248 .040402

JC10 Rec'd PCT/PTO 14 DEC 2001

FORM PTO-1390 (REV. 9-2001) U.S. DEPARTMENT OF COMMERCE PATENT AND TRADEMARK OFFICE

TRANSMITTAL LETTER TO THE UNITED STATES DESIGNATED/ELECTED OFFICE (DO/EO/US) CONCERNING A FILING UNDER 35 U.S.C. 371

ATTORNEY'S DOCKET NUMBER: EX99-004C-US

U.S. APPLICATION NO. (If known, see 37 CFR 1.5): 10/018248

INTERNATIONAL APPLICATION NO.	INTERNATIONAL FILING DATE	PRIORITY DATE CLAIMED
PCT/US00/15880	June 8, 2000	June 14, 1999

TITLE OF INVENTION: Animal Models and Methods for Analysis of Lipid Metabolism and Screening of Pharmaceutical and Pesticidal Agents that Modulate Lipid Metabolism

APPLICANT(S) FOR DO/EO/US: Costa et al.

Applicant herewith submits to the United States Designated/Elected Office (DO/EO/US) the following items and other information:

1. ☑ This is a **FIRST** submission of items concerning a filing under 35 U.S.C. 371.
2. ☐ This is a **SECOND** or **SUBSEQUENT** submission of items concerning a filing under 35 U.S.C. 371.
3. ☑ This is an express request to begin national examination procedures (35 U.S.C. 371(f)). The submission must include items (5), (6), (9) and (21) indicated below.
4. ☐ The US has been elected by the expiration of 19 months from the priority date (Article 31).
5. ☑ A copy of the International Application as filed (35 U.S.C. 371(c)(2))
   a. ☐ is attached hereto (required only if not communicated by the International Bureau).
   b. ☐ has been communicated by the International Bureau.
   c. ☑ is not required, as the application was filed in the United States Receiving Office (RO/US).
6. ☑ An English language translation of the International Application as filed (35 U.S.C. 371(c)(2)).
   a. ☐ is attached hereto.
   b. ☑ has been previously submitted under 35 U.S.C. 154(d)(4).
7. ☐ Amendments to the claims of the International Aplication under PCT Article 19 (35 U.S.C. 371(c)(3))
   a. ☐ are attached hereto (required only if not communicated by the International Bureau).
   b. ☐ have been communicated by the International Bureau.
   c. ☐ have not been made; however, the time limit for making such amendments has NOT expired.
   d. ☐ have not been made and will not be made.
8. ☐ An English language translation of the amendments to the claims under PCT Article 19 (35 U.S.C. 371 (c)(3)).
9. ☑ An oath or declaration of the inventor(s) (35 U.S.C. 371(c)(4)). (unsigned)
10. ☐ An English lanugage translation of the annexes of the International Preliminary Examination Report under PCT Article 36 (35 U.S.C. 371(c)(5)).

Items 11 to 20 below concern document(s) or information included:

11. ☐ An Information Disclosure Statement under 37 CFR 1.97 and 1.98.
12. ☐ An assignment document for recording. A separate cover sheet in compliance with 37 CFR 3.28 and 3.31 is included.
13. ☐ A FIRST preliminary amendment.
14. ☐ A SECOND or SUBSEQUENT preliminary amendment.
15. ☐ A substitute specification.
16. ☐ A change of power of attorney and/or address letter.
17. ☐ A computer-readable form of the sequence listing in accordance with PCT Rule 13ter.2 and 35 U.S.C. 1.821 - 1.825.
18. ☐ A second copy of the published international application under 35 U.S.C. 154(d)(4).
19. ☐ A second copy of the English language translation of the international application under 35 U.S.C. 154(d)(4).
20. ☐ Other items or information:

JC05 Rec'd PCT/PTO 14 DEC 2001

U.S. APPLICATION NO. (if known, see 37 CFR 1.5)	INTERNATIONAL APPLICATION NO.	ATTORNEY'S DOCKET NUMBER
10/018248	PCT/US00/15880	EX99-004C-US

21. ☑ The following fees are submitted:	CALCULATIONS PTO USE ONLY	
BASIC NATIONAL FEE (37 CFR 1.492 (a) (1) - (5)):		
Neither international preliminary examination fee (37 CFR 1.482) nor international search fee (37 CFR 1.445(a)(2)) paid to USPTO and International Search Report not prepared by the EPO or JPO **$1040.00**		
International preliminary examination fee (37 CFR 1.482) not paid to USPTO but International Search Report prepared by the EPO or JPO **$890.00**		
International preliminary examination fee (37 CFR 1.482) not paid to USPTO but international search fee (37 CFR 1.445(a)(2)) paid to USPTO **$740.00**		
International preliminary examination fee (37 CFR 1.482) paid to USPTO but all claims did not satisfy provisions of PCT Article 33(1)-(4) **$710.00**		
International preliminary examination fee (37 CFR 1.482) paid to USPTO and all claims satisfied provisions of PCT Article 33(1)-(4) **$100.00**		
ENTER APPROPRIATE BASIC FEE AMOUNT =	$ 710.00	
Surcharge of **$130.00** for furnishing the oath or declaration later than ☐ 20 ☐ 30 months from the earliest claimed priority date (37 CFR 1.492(e)).	$	

CLAIMS	NUMBER FILED	NUMBER EXTRA	RATE	
Total claims	31 - 20 =	11	x **$18.00**	$ 198.00
Independent claims	4 - 3 =	1	x **$84.00**	$ 84.00
MULTIPLE DEPENDENT CLAIM(S) (if applicable)			+ **$280.00**	$ Ø
TOTAL OF ABOVE CALCULATIONS =				$ 992.00
☐ Applicant claims small entity status. See 37 CFR 1.27. The fees indicated above are reduced by 1/2.			+	$ Ø
SUBTOTAL =				$
Processing fee of **$130.00** for furnishing the English translation later than ☐ 20 ☐ 30 months from the earliest claimed priority date (37 CFR 1.492(f)).				$ Ø
TOTAL NATIONAL FEE =				$
Fee for recording the enclosed assignment (37 CFR 1.21(h)). The assignment must be accompanied by an appropriate cover sheet (37 CFR 3.28, 3.31). **$40.00** per property +				$ Ø
TOTAL FEES ENCLOSED =				$ 992.00
			Amount to be refunded:	$
			charged:	$

a. ☐ A check in the amount of $ _____ to cover the above fees is enclosed.

b. ☑ Please charge my Deposit Account No. 50-1108 in the amount of $ 992.00 to cover the above fees. A duplicate copy of this sheet is enclosed.

c. ☑ The Commissioner is hereby authorized to charge any additional fees which may be required, or credit any overpayment to Deposit Account No. 50-1108. A duplicate copy of this sheet is enclosed.

d. ☐ Fees are to be charged to a credit card. **WARNING:** Information on this form may become public. **Credit card information should not be included on this form.** Provide credit card information and authorization on PTO-2038.

NOTE: Where an appropriate time limit under 37 CFR 1.494 or 1.495 has not been met, a petition to revive (37 CFR 1.137 (a) or (b)) must be filed and granted to restore the application to pending status.

SEND ALL CORRESPONDENCE TO:

Jan P. Brunelle
Exelixis, Inc.
170 Harbor Way
P.O. Box 511
South San Francisco, CA 94083-0511

Jan P. Brunelle
SIGNATURE

Jan P. Brunelle
NAME

35,081
REGISTRATION NUMBER

ANIMAL MODELS AND METHODS FOR ANALYSIS OF LIPID METABOLISM AND SCREENING OF PHARMACEUTICAL AND PESTICIDAL AGENTS THAT MODULATE LIPID METABOLISM

BACKGROUND OF THE INVENTION

There is much interest within the pharmaceutical industry to understand the mechanisms involved in cholesterol synthesis and metabolism, particularly on the molecular level, so that blood cholesterol lowering drugs can be developed for the treatment or prevention of atherosclerosis. There is further interest in understanding the molecular mechanisms that connect lipid defects and insulin resistance. Hyperlipidemia and elevation of free fatty acid levels correlate with "Metabolic Syndrome," defined as the linkage between several diseases, including obesity and insulin resistance, which often occur in the same patients and which are major risk factors for development of Type 2 diabetes and cardiovascular disease. Current research suggests that the control of lipid levels, in addition to glucose levels, may be required to treat Type 2 Diabetes, heart disease, and other manifestations of Metabolic Syndrome (Santomauro AT *et al.*, Diabetes (1999) 48:1836-1841).

Recent advances have been made in understanding some of the mechanisms involved in mammalian lipid metabolism. A key component is the sterol regulatory element binding protein (SREBP) pathway. SREBPs are transcription factors that activate genes involved in cholesterol and fatty acid synthesis and transport. SREBP is the major mediator of insulin action in the liver, and alterations in expression and function of SREBPs have been described in obese and insulin resistant patients or animal models (Shimomura I *et al.*, PNAS (1999) 96:13656-61; Shimomura I *et al.*, Journal of Biological Chemistry (1999) 274:30028-32). SREBPs are also implicated in the process of fat cell differentiation and adipose cell gene expression, particularly as transcription factors that can promote adipogenesis in a dominant fashion (reviewed by Spiegelman *et al.*, Cell (1996) 87:377-389). SREBP function is regulated by intracellular levels of sterols or polyunsaturated fatty acids (PUFAs) (Xu J. *et al.*, J. Biol. Chem. (1999) 274:23577-23583).

In high sterol or PUFA conditions, SREBPs are retained as membrane-bound protein precursors that are kept inactive by virtue of being attached to the nuclear envelope and endoplasmic reticulum (ER) and therefore, excluded from the nucleus. An SREBP in its membrane-bound form has large N-terminal and C-terminal

segments facing the cytoplasm and a short loop projecting into the lumen of the organelle. The N-terminal domain is a transcription factor of the basic-helix-loop-helix-leucine zipper (bHLH-Zip) family, and contains an "acid blob" typical of many transcriptional activators (Brown and Goldstein, Cell (1997) 89:331-340). The N-terminal acid blob is followed by a basic helix-loop-helix/leucine zipper domain (bHLH-Zip) similar to those found in many other DNA-binding transcriptional regulators.

In low sterol conditions, the acid blob/bHLH-Zip domain of SREBP is released from the membrane after which it is rapidly translocated into the nucleus and binds specific DNA sequences to activate transcription. Two sequential proteolytic cleavages are involved. A first protease, referred to as the site 1 protease (S1P) cleaves SREBP at approximately the middle of the lumenal loop. S1P has been cloned from Chinese hamster ovary (CHO) cells (GI (GenBank Identifier No. (hereinafter "GI") 3892203) and a human cell line (GI4506774) (Sakai et al., J. Biol. Chem (1998) 273:5785-5793).

After cleavage at site 1, a second protease (the site 2 protease, S2P) cleaves the N-terminal fragment and releases the mature N-terminal domain into the cytosol, from which it rapidly enters the nucleus, apparently with a portion of the transmembrane domain still attached at the C-terminus. Mature, transcriptionally active SREBP is rapidly degraded in a proteosome-dependent process. This combination of proteolytic processing and rapid turnover allows the SREBP system to rapidly respond to changes in cellular membrane components. S2P homologues have been identified in both vertebrates and invertebrates and have been cloned from human cells and hamster cells (Rawson *et al.*, Molec Cell (1997) 1:47-57).

A third component of the processing system for SREBPs is called SREBP Cleavage Activating Protein (SCAP). SCAP is a large transmembrane protein that activates S1P in low-sterol conditions (Hua *et al.*, Cell (1996) 87:415-426). To date, the SREBP pathway has been studied primarily using mammalian cell culture, by the isolation of mutant cells that are defective in regulation of cholesterol metabolism or intracellular cholesterol trafficking. The mutants can then serve as hosts for cloning genes by functional complementation. This has led to the molecular cloning of the S1P, S2P and SCAP genes (Rawson et al., supra; Hua et al., supra; Goldstein et al., US Pat. Nos. 5,527,690 and 5,891,631 and PCT Application No. WO00/09677).

Some SREBP pathway genes have been identified in invertebrates. The isolation of a *Drosophila* SREBP, referred to as "HLH106" (GI079656) has been described (Theopold et al., Proc. Natl. Acad. Sci., USA, (1996) 93(3):1195-1199). An expressed sequence tag (EST) from *Caenorhabditis elegans* (*C. elegans)* which has

homology to S2P is described by Rawson et al., supra and is listed in GenBank (GI1559384). Additionally, GenBank has listed a protein predicted from the *C. elegans* genome as having HMG-CoA reductase homology (GI3875380).

The present invention discloses invertebrate SREBP pathway nucleic acids and proteins and methods of use.

SUMMARY OF THE INVENTION

The use of invertebrate model organism genetics can greatly facilitate the elucidation of biochemical pathways, and the identification of molecules that can modulate such pathways. Accordingly, it is an object of the invention to provide invertebrate nucleic acids and polypeptides involved in the SREBP pathway. It is also an object of the invention to provide invertebrate model organisms, including novel mutant phenotypes, for the study of lipid metabolism in general, and more particularly, for the elucidation of the SREBP pathway. It is a further object of the invention to provide methods for screening molecules that modulate lipid metabolism and/or the function of genes and proteins involved in the SREBP pathway.

These and other objects are provided by flies and nematodes that have been genetically modified to express or misexpress an SREBP pathway gene, for example using transposon mutagenesis, RNA interference, chemical mutagenesis, or other genetic techniques. In certain embodiments, expression of the SREBP pathway protein is driven by a heterologous promoter that is tissue-specific, developmentally-specific, or inducible, so that the effects of the expression or mis-expression can be observed in specific tissues, at certain developmental stages, or at specified times, respectively. Additionally, the SREBP pathway protein may be linked to one or more selectable markers that allows detection of expression. Typically, the expression of the SREBP pathway protein results in an identifiable phenotype. In the case of nematodes, the invention provides novel methods for the in vivo measurement of lipid content using BODIPY-fatty acid conjugates. The animal models can be used in genetic screens to identify other genes involved in lipid metabolism. They can also be used for screening small molecule libraries directly on whole organisms for possible therapeutic or pesticide use.

The invention also provides novel isolated nucleic acids (SEQ ID NOs:1, 3, 5 and 7) and the SREBP pathway proteins encoded thereby (SEQ ID NOs:2, 4, 6, and 8 respectively), as well as derivatives and fragments thereof. Methods are provided for constructing vectors containing the isolated nucleic acids. Such vectors can be used for making the animal models of the invention. They can also be introduced into host cells to be used for a variety of purposes including two-hybrid screening assays,

production of SREBP pathway proteins, screening small molecules that affect lipid synthesis or metabolism, etc.

BRIEF DESCRIPTION OF THE DRAWINGS

Figs. 1A and 1B depict the inactive, membrane-bound form of SREBP (Fig. 1A) and the two-step proteolytic cleavage that activates SREBP in low sterol conditions (Fig. 1B).

Fig. 2 depicts the presumed interactions between SREBP, SCAP, S1P and S2P in the SREBP processing complex.

DETAILED DESCRIPTION OF THE INVENTION

The use of invertebrate model organism genetics and related technologies can greatly facilitate the elucidation of biological pathways (Scangos, Nat. Biotechnol. (1997) 15:1220-1221; Margolis and Duyk, *supra*). Of particular use are the insect and nematode model organism, *Drosophila melanogaster* (hereinafter referred to generally as "*Drosophila*") and *C. elegans*. Novel SREBP pathway nucleic acids, from *Drosophila* and *C. elegans*, and their encoded proteins are identified herein. As used in this description, the term "SREBP pathway nucleic acid" refers to a nucleic acid that encodes any one of SREBP, SCAP, S1P, and S2P. The newly identified SREBP pathway nucleic acids have led to the discovery of several mutant phenotypes that can be used to study the pathways involved in lipid and fatty acid metabolism. The use of invertebrate model organisms, such as *Drosophila* and *C. elegans,* for analyzing the expression and mis-expression of SREBP pathway proteins has great advantages over the traditional approach of using mammalian cell culture due to the ability to rapidly carry out large-scale, systematic genetic screens. These can identify other components involved in the synthesis, activation, control, and turnover of SREBP pathway proteins. Additionally, model organisms or cultured cells that have been genetically engineered to misexpress SREBP pathway genes can be used to screen candidate compounds or small molecule libraries for their ability to modulate the expression or activity of SREBP pathway proteins. They are therefore useful in the identification of new drug targets, therapeutic agents, diagnostics and prognostics for the treatment of disorders associated with lipid metabolism. Additionally, use of these invertebrate model organisms could lead to the identification and validation of pesticide targets directed to components of the SREBP pathway.

Nucleic acids of the SREBP pathway

The present invention provides a nucleic acid sequence (SEQ ID NO:1) that was isolated from *C. elegans* and encodes an SREBP homologue referred to herein as "ceSREBP". The invention also provides nucleic acid sequences that were isolated from *Drosophila* and encode homologues of S2P (dS2P; SEQ ID NO:3), SCAP (dSCAP; SEQ ID NO:5) and S1P (dS1P; SEQ ID NO:7). In addition to the fragments and derivatives of SEQ ID NOs 1, 3, 5, and 7, as described in detail below, the invention includes the reverse complements thereof. Also, the subject nucleic acid sequences, derivatives and fragments thereof may be RNA molecules comprising the nucleotide sequence of any one of SEQ ID NOs 1, 3, 5 and 7 (or derivative or fragment thereof) wherein the base U (uracil) is substituted for the base T (thymine). The DNA and RNA sequences of the invention can be single- or double-stranded. Thus, the term "nucleic acid sequence", as used herein, includes the reverse complement, RNA equivalent, DNA or RNA double-stranded sequences, and DNA/RNA hybrids of the sequence being described, unless otherwise indicated explicitly or by context.

Fragments of these sequences can be used for a variety of purposes. Interfering RNA (RNAi) fragments, particularly double-stranded (ds) RNAi, can be used to generate loss-of-function phenotypes. SREBP pathway nucleic acid fragments are also useful as nucleic acid hybridization probes and replication/amplification primers. Certain "antisense" fragments, i.e. that are reverse complements of portions of the coding sequence of SEQ ID NOs: 1, 3, 5, and 7 have utility in inhibiting the function of SREBP pathway proteins. The fragments are of length sufficient to specifically hybridize with the corresponding SEQ ID NOs:1, 3, 5, and 7. The fragments consist of or comprise at least 12, preferably at least 24, more preferably at least 36, and more preferably at least 96 contiguous nucleotides of SEQ ID NOs: 1, 3, 5, and 7. When the fragments are flanked by other nucleic acid sequences, the total length of the combined nucleic acid sequence is less than 15 kilobases (kb), preferably less than 10 kb or less than 5kb, and more preferably less than 2 kb.

Preferred fragments of ceSREBP (SEQ ID NO:1) contain approximately residues 1090 to 1290, which encode the bHLH-Zip domain.

Preferred fragments of dS2P (SEQ ID NO:3) include those having at least 1226 contiguous nucleotides of SEQ ID NO:3, and more preferably at least 1231 nucleotides.

Preferred fragments of dSCAP (SEQ ID NO:5) encode the WD repeats, which are located at approximately nucleotides 2509-2617, 3080-3196, 3208-3325, and 3337-3445.

Preferred fragments of dS1P (SEQ ID NO:7) encode the extracellular or intracellular domains, which are located at approximately nucleotides 131-1149, 1214-1434, 1499-1560, and 1625-3040. Other preferred fragments consist or comprise at least 12 contiguous nucleotides, preferably at least 37 contiguous nucleotides, and more preferably at least 62 contiguous nucleotides of nucleotides 2015-2546 of SEQ ID NO:7.

Additionally, fragments of any of the foregoing sequences that are double-stranded RNA (dsRNA) molecules have utility in RNA interference (RNAi) studies, as described in more detail below, where model organisms exhibiting loss-of-function phenotype are generated. Typically, dsRNA molecules for RNAi studies are from about 200 to 2000 bp, and are preferably 600-900 bp in size.

The subject nucleic acid sequences may consist solely of SEQ ID NOs:1, 3, 5, and 7 or fragments thereof. Alternatively, the subject nucleic acid sequences and fragments thereof may be joined to other components such as labels, peptides, agents that facilitate transport across cell membranes, hybridization-triggered cleavage agents or intercalating agents. The subject nucleic acid sequences and fragments thereof may also be joined to other nucleic acid sequences (*i.e.* they may comprise part of larger sequences) and are of synthetic/non-natural sequences. They may be isolated and/or are purified and thus unaccompanied by at least some of the material with which they associate in the natural state. Preferably, the isolated nucleic acids constitute at least about 0.5%, and more preferably at least about 5%, by weight, of the total nucleic acid present in a given fraction, and are preferably recombinant, meaning that they comprise a non-natural sequence or a natural sequence joined to nucleotide(s) other than that which it is joined to on a natural chromosome.

Derivative SREBP pathway nucleic acid sequences include sequences that hybridize to the nucleic acid sequence of SEQ ID NO:1, 3, 5 or 7 under stringency conditions such that the hybridizing derivative nucleic acid is related to the subject nucleic acid by a certain degree of sequence identity. A nucleic acid molecule is "hybridizable" to another nucleic acid molecule, such as a cDNA, genomic DNA, or RNA, when a single stranded form of the nucleic acid molecule can anneal to the other nucleic acid molecule. Stringency of hybridization refers to conditions under which nucleic acids are hybridizable. The degree of stringency can be controlled by temperature, ionic strength, pH, and the presence of denaturing agents such as formamide during hybridization and washing. As used herein, the term "stringent hybridization conditions" are those normally used by one of skill in the art to establish at least a 90% sequence identity between complementary pieces of DNA or DNA and RNA. "Moderately stringent hybridization conditions" are used to find derivatives

having at least 70% sequence identity. Finally, "low-stringency hybridization conditions" are used to isolate derivative nucleic acid molecules that share at least about 50% sequence identity with the subject nucleic acid sequence.

The ultimate hybridization stringency reflects both the actual hybridization conditions as well as the washing conditions following the hybridization, and it is well known in the art how to vary the conditions to obtain the desired result. Conditions routinely used are set out in readily available procedure texts (*e.g.*, Current Protocol in Molecular Biology, Vol. 1, Chap. 2.10, John Wiley & Sons, Publishers (1994); Sambrook et al., Molecular Cloning, Cold Spring Harbor (1989)). A preferred derivative nucleic acid is capable of hybridizing to SEQ ID NO:1, 3, 5, or 7 under stringent hybridization conditions that comprise: prehybridization of filters containing nucleic acid for 8 hours to overnight at 65° C in a solution comprising 6X single strength citrate (SSC) (1X SSC is 0.15 M NaCl, 0.015 M Na citrate; pH 7.0), 5X Denhardt's solution, 0.05% sodium pyrophosphate and 100 µg/ml herring sperm DNA; hybridization for 18-20 hours at 65° C in a solution containing 6X SSC, 1X Denhardt's solution, 100 µg/ml yeast tRNA and 0.05% sodium pyrophosphate; and washing of filters at 65° C for 1 h in a solution containing 0.2X SSC and 0.1% SDS (sodium dodecyl sulfate).

Derivative nucleic acid sequences that have at least about 70% sequence identity with SEQ ID NOs:1, 3, 5 or 7 are capable of hybridizing to SEQ ID NOs:1, 3, 5, and 7, respectively, under moderately stringent conditions that comprise: pretreatment of filters containing nucleic acid for 6 h at 40° C in a solution containing 35% formamide, 5X SSC, 50 mM Tris-HCl (pH 7.5), 5 mM EDTA, 0.1% PVP, 0.1% Ficoll, 1% BSA, and 500 µg/ml denatured salmon sperm DNA; hybridization for 18-20 h at 40° C in a solution containing 35% formamide, 5X SSC, 50 mM Tris-HCl (pH 7.5), 5 mM EDTA, 0.02% PVP, 0.02% Ficoll, 0.2% BSA, 100 µg/ml salmon sperm DNA, and 10% (wt/vol) dextran sulfate; followed by washing twice for 1 hour at 55° C in a solution containing 2X SSC and 0.1% SDS.

Other preferred derivative nucleic acid sequences are capable of hybridizing to SEQ ID NOs:1, 3, 5, or 7 under low stringency conditions that comprise: incubation for 8 hours to overnight at 37° C in a solution comprising 20% formamide, 5 x SSC, 50 mM sodium phosphate (pH 7.6), 5X Denhardt's solution, 10% dextran sulfate, and 20 µg/ml denatured sheared salmon sperm DNA; hybridization in the same buffer for 18 to 20 hours; and washing of filters in 1 x SSC at about 37° C for 1 hour.

As used herein, "percent (%) nucleic acid sequence identity" with respect to a subject sequence, or a specified portion of a subject sequence, is defined as the percentage of nucleotides in the candidate derivative nucleic acid sequence identical

with the nucleotides in the subject sequence (or specified portion thereof), after aligning the sequences and introducing gaps, if necessary to achieve the maximum percent sequence identity, as generated by the program WU-BLAST-2.0a19 (Altschul *et al.*, J. Mol. Biol. (1997) 215:403-410; http://blast.wustl.edu/blast/README.html; hereinafter referred to generally as "BLAST") with all the search parameters set to default values. A percent (%) nucleic acid sequence identity value is determined by the number of matching identical nucleotides divided by the sequence length for which the percent identity is being reported.

Derivative SREBP pathway nucleic acid sequences usually have at least 70% sequence identity, preferably at least 80% sequence identity, more preferably at least 85% sequence identity, still more preferably at least 90% sequence identity, and most preferably at least 95% sequence identity with SEQ ID NO:1, 3, 5 or 7, or domain-encoding regions thereof.

Preferred derivatives of SEQ ID NO:7 comprise a nucleotide sequence having at least 53% sequence identity, and preferably at least 60% sequence identity with any contiguous 125 or 275 bases of nucleotides 2015-2546 of SEQ ID NO:7, or the reverse complement thereof.

In one preferred embodiment, the derivative nucleic acids encode polypeptides comprising SREBP pathway amino acid sequence of SEQ ID NOs:2, 4, 6, and 8 or fragments or derivatives thereof as described further below under the subheading "SREBP pathway proteins". A derivative SREBP pathway nucleic acid sequence, or fragment thereof, may comprise 100% sequence identity with SEQ ID NO:1, 3, 5, or 7 but be a derivative thereof in the sense that it has one or more modifications at the base or sugar moiety, or phosphate backbone. Examples of modifications are well known in the art (Bailey, Ullmann's Encyclopedia of Industrial Chemistry (1998), 6th ed. Wiley and Sons). Such derivatives may be used to provide modified stability or any other desired property.

Another type of derivative of the subject nucleic acid sequences includes the corresponding humanized sequences. A humanized nucleic acid sequence is one in which one or more codons has been substituted with a codon that is more commonly used in human genes. Preferably, a sufficient number of codons have been substituted such that a higher level expression is achieved in mammalian cells than what would otherwise be achieved without the substitutions. Codons that are more commonly used in human genes are known (Wada *et al.*, Nucleic Acids Research (1990) 18(Suppl.):2367-2411). Also, a detailed discussion of the humanization of nucleic acid sequences is provided in U.S. Pat. No. 5,874,304 to Zolotukhin *et al.* Similarly, other nucleic acid derivatives can be generated with codon usage optimized for

expression in other organisms, such as yeasts, bacteria, and plants, where it is desired to engineer the expression of SREBP pathway proteins by using specific codons chosen according to the preferred codons used in highly expressed genes in each organism. More specific embodiments of preferred SREBP pathway protein fragments and derivatives are further discussed below.

Nucleic acid encoding the amino acid sequence of any one of SEQ ID NO:2, 4, 6, or 8, or a fragment or derivative thereof, may be obtained from an appropriate cDNA library prepared from any eukaryotic species that encodes SREBP pathway proteins such as vertebrates, preferably mammalian (*e.g.* primate, porcine, bovine, feline, equine, and canine species, *etc.*) and invertebrates, such as arthropods, particularly insects species (preferably *Drosophila*), acarids, crustacea, molluscs, nematodes (preferably *C. elegans*), and other worms. An expression library can be constructed using known methods. For example, mRNA can be isolated to make cDNA that is ligated into a suitable expression vector for expression in a host cell into which it is introduced. Various screening assays can then be used to select for the gene or gene product (*e.g.* oligonucleotides of at least about 20 to 80 bases designed to identify the gene of interest, or labeled antibodies that specifically bind to the gene product). The gene and/or gene product can then be recovered from the host cell using known techniques.

Polymerase chain reaction (PCR) can also be used to isolate nucleic acids of the SREBP pathway where oligonucleotide primers representing fragmentary sequences of interest amplify RNA or DNA sequences from a source such as a genomic or cDNA library (as described by Sambrook *et al.*, *supra*). Additionally, degenerate primers for amplifying homologs from any species of interest may be used. Once a PCR product of appropriate size and sequence is obtained, it may be cloned and sequenced by standard techniques, and utilized as a probe to isolate a complete cDNA or genomic clone.

Fragmentary sequences of SEQ ID NOs 1, 3, 5 and 7 nucleic acids and derivatives may be synthesized by known methods. For example, oligonucleotides may be synthesized using an automated DNA synthesizer available from commercial suppliers (*e.g.* Biosearch, Novato, CA; Perkin-Elmer Applied Biosystems, Foster City, CA). Antisense RNA sequences can be produced intracellularly by transcription from an exogenous sequence, *e.g.* from vectors that contain antisense SREBP pathway nucleic acid sequences. Newly generated sequences may be identified and isolated using standard methods.

An isolated SREBP pathway nucleic acid sequence can be inserted into any appropriate cloning vector, for example bacteriophages such as lambda derivatives, or

plasmids such as PBR322, pUC plasmid derivatives and the Bluescript vector (Stratagene, San Diego, CA). Recombinant molecules can be introduced into host cells via transformation, transfection, infection, electroporation, *etc.*, or into a transgenic animal such as a fly. The transformed cells can be cultured to generate large quantities of the SREBP pathway nucleic acid. Suitable methods for isolating and producing the subject nucleic acid sequences are well known in the art (Sambrook *et al.*, *supra*; DNA Cloning: A Practical Approach, Vol. 1, 2, 3, 4, (1995) Glover, ed., MRL Press, Ltd., Oxford, U.K.).

The nucleotide sequence encoding an SREBP pathway protein, or a fragment or derivative thereof, can be inserted into any appropriate expression vector for the transcription and translation of the inserted protein-coding sequence. Alternatively, the native SREBP pathway gene and/or its flanking regions can supply the necessary transcriptional and translational signals. A variety of host-vector systems may be utilized to express the protein-coding sequence. These include mammalian cell systems infected with virus (*e.g.* vaccinia virus, adenovirus, *etc.*), insect cell systems infected with virus (*e.g.* baculovirus), microorganisms such as yeast containing yeast vectors, or bacteria transformed with bacteriophage, DNA, plasmid DNA, or cosmid DNA. Expression of a SREBP pathway protein may be controlled by a suitable promoter/enhancer element. In addition, a host cell strain may be selected which modulates the expression of the inserted sequences, or modifies and processes the gene product in the specific fashion desired.

To detect expression of the SREBP pathway gene product, the expression vector can comprise a promoter operably linked to an SREBP pathway nucleic acid, one or more origins of replication, and, one or more selectable markers (*e.g.* thymidine kinase activity, resistance to antibiotics, *etc.*). Alternatively, recombinant expression vectors can be identified by assaying for the expression of the SREBP pathway gene product based on the physical or functional properties of the SREBP pathway protein in *in vitro* assay systems (*e.g.* immunoassays).

The SREBP pathway protein, fragment, or derivative may be optionally expressed as a fusion, or chimeric protein product (*i.e.*, joined via a peptide bond to a heterologous protein sequence of a different protein). A chimeric product can be made by ligating the appropriate nucleic acid sequences encoding the desired amino acid sequences to each other in the proper coding frame using standard methods and expressing the chimeric product. A chimeric product may also be made by protein synthetic techniques, *e.g.* by use of a peptide synthesizer.

Once a recombinant that expresses the SREBP pathway gene sequence is identified, the gene product can be isolated and purified using standard methods (*e.g.*

ion exchange, affinity, and gel exclusion chromatography; centrifugation; differential solubility; electrophoresis). The amino acid sequence of the protein can be deduced from the nucleotide sequence of the chimeric gene contained in the recombinant and can thus be synthesized by standard chemical methods (Hunkapiller *et al.*, Nature (1984) 310:105-111). Alternatively, native SREBP pathway proteins can be purified from natural sources, by standard methods (*e.g.* immunoaffinity purification).

SREBP pathway proteins

SREBP pathway proteins of the invention comprise or consist of an amino acid sequence of any one of SEQ ID NO: 2, 4, 6 or 8, or fragments or derivatives thereof. Compositions comprising these proteins may consist essentially of the SREBP pathway protein, fragments, or derivatives, or may comprise additional components (*e.g.* pharmaceutically acceptable carriers or excipients, culture media, *etc.*).

SREPB pathway protein derivatives typically share a certain degree of sequence identity or sequence similarity with any one of SEQ ID NOs: 2, 4, 6, and 8, or a fragment thereof. As used herein, "percent (%) amino acid sequence identity," with respect to a subject sequence, or a specified portion of a subject sequence, is defined as the percentage of amino acids in the candidate derivative amino acid sequence identical to the amino acid in the subject sequence (or specified portion thereof), after aligning the sequences and introducing gaps as necessary to achieve the maximum percent sequence identity as generated by BLAST (Altschul *et al.*, *supra*) using the same parameters discussed above for derivative nucleic acid sequences. A % amino acid sequence identity value is determined by the number of matching identical amino acids divided by the sequence length for which the percent identity is being reported. "Percent (%) amino acid sequence similarity" is determined by doing the same calculation as for determining % amino acid sequence identity, but including conservative amino acid substitutions in addition to identical amino acids in the computation. A conservative amino acid substitution is one in which an amino acid is substituted for another amino acid having similar properties such that the folding or activity of the protein is not significantly affected. Aromatic amino acids that can be substituted for each other are phenylalanine, tryptophan, and tyrosine. Interchangeable hydrophobic amino acids are leucine, isoleucine, methionine, and valine. Interchangeable polar amino acids are glutamine and asparagine. Interchangeable basic amino acids are arginine, lysine and histidine. Interchangeable acidic amino acids are aspartic acid and glutamic acid; and interchangeable small amino acids are alanine, serine, cysteine, threonine, and glycine.

In one preferred embodiment, an SREBP pathway protein derivative shares at least 75% sequence identity or similarity, preferably at least 80%, more preferably at least 85%, still more preferably at least 90% and most preferably at least 95% sequence identity or similarity with a contiguous stretch of at least 25 amino acids, preferably at least 50 amino acids, more preferably at least 100 amino acids, and in some cases, the entire length of any of SEQ ID NO:2, 4, 6, or 8. Other preferred derivatives of ceSREBP consist of or comprise an amino acid sequence that shares at least 75% similarity, preferably at least 80% similarity, and more preferably, at least 85% similarity with amino acid residues 335-428 of SEQ ID NO:2. Preferably, such derivatives share antigenicity with amino acid residues 335-428 of SEQ ID NO:2.

Other preferred derivatives of dSCAP consist of or comprise an amino acid sequence that shares at least 75% similarity, preferably at least 80% similarity, and more preferably, at least 85% similarity with amino acid residues 812-848, 1005-1041, 1045-1084, and 1088-1124 of SEQ ID NO:6, which constitute the WD repeats. Preferably, such derivatives share antigenicity with amino acid residues 812-848, 1005-1041, 1045-1084, and 1088-1124 of SEQ ID NO:6.

In another embodiment, the dS1P protein derivative may consist of or comprise a sequence that shares 100% similarity with any contiguous stretch of at least 33 amino acids, preferably at least 35 amino acids, more preferably at least 38 amino acids, and most preferably at least 43 amino acids of SEQ ID NO:8. Preferred derivatives of dS1P consist of or comprise an amino acid sequence that has at least 70%, preferably at least 80%, more preferably at least 85%, still more preferably at least 90%, and most preferably at least 95% sequence identity or sequence similarity with any of amino acid residues 22-362, 384-457, and 521-993, which are extracellular or intracellular domains of dS1P. Another preferred derivative of dS1P protein consists of or comprises a sequence of at least 10 amino acids that share 100% similarity with an equivalent number of contiguous amino acids of residues 653-828 of SEQ ID NO:8.

Preferred fragments of dS1P proteins consist or comprise at least 16, preferably at least 18, more preferably at least 21, and most preferably at least 26 contiguous amino acids of SEQ ID NO:2. Other preferred fragments include any 10 contiguous amino acids, preferably any 20 contiguous amino acids, and more preferably any 60 contiguous amino acids of residues 653-828 of SEQ ID NO:2.

The fragment or derivative of an SREBP pathway protein is preferably "functionally active" meaning that the SREBP pathway protein derivative or fragment exhibits one or more functional activities associated with a full-length, wild-type SREBP pathway protein comprising the amino acid sequence of any one of SEQ ID

NO:2, 4, 6 or 8. As one example, a fragment or derivative may have antigenicity such that it can be used in immunoassays, for immunization, for inhibition of SREBP pathway activity, *etc.*, as discussed further below regarding generation of antibodies to SREBP pathway proteins. As another example, a fragment or derivative of SREBP may be considered functionally active if it binds a regulatory DNA element of an appropriate target gene such as the SRE-1 sequence. S1P may be considered functionally active if it cleaves SREBP at site 1 (as depicted in Fig. 1B). S2P may be considered functionally active if it cleaves SREBP at site 2. A fragment or derivative of SCAP may be considered functionally active if it binds to the C-terminal, regulatory domain of SREBP. The functional activity of SREBP pathway proteins, derivatives and fragments can be assayed by various methods known to one skilled in the art (Current Protocols in Protein Science (1998) Coligan *et al.*, *eds.*, John Wiley & Sons, Inc., Somerset, New Jersey). In a preferred method, which is described in detail below, a model organism, such as an insect (*e.g., Drosophila*) or worm (*e.g., C. elegans*), or other model system, is used in genetic studies to assess the phenotypic effect of a fragment or derivative (i.e. mutant).

The SREBP pathway derivatives can be produced by various methods known in the art. The manipulations that result in their production can occur at the gene or protein level. For example, a cloned SREBP pathway gene sequence can be cleaved at appropriate sites with restriction endonuclease(s) (Wells *et al.*, Philos. Trans. R. Soc. London SerA (1986) 317:415), followed by further enzymatic modification if desired, isolated, and ligated *in vitro*, and expressed to produce the desired derivative. Alternatively, an SREBP pathway gene can be mutated *in vitro* or *in vivo*, to create and/or destroy translation, initiation, and/or termination sequences, or to create variations in coding regions and/or to form new restriction endonuclease sites or to destroy preexisting ones, to facilitate further *in vitro* modification. A variety of mutagenesis techniques are known in the art such as chemical mutagenesis, *in vitro* site-directed mutagenesis (Carter *et al.*, Nucl. Acids Res. (1986) 13:4331), use of TAB® linkers (available from Pharmacia and Upjohn, Kalamazoo, MI), *etc.*

At the protein level, manipulations include post translational modification, *e.g.* glycosylation, acetylation, phosphorylation, amidation, derivatization by known protecting/blocking groups, proteolytic cleavage, linkage to an antibody molecule or other cellular ligand, *etc.* Any of numerous chemical modifications may be carried out by known technique (*e.g.* specific chemical cleavage by cyanogen bromide, trypsin, chymotrypsin, papain, V8 protease, $NaBH_4$, acetylation, formylation, oxidation, reduction, metabolic synthesis in the presence of tunicamycin, *etc.*). Derivative proteins can also be chemically synthesized by use of a peptide synthesizer,

for example to introduce nonclassical amino acids or chemical amino acid analogs as substitutions or additions into the SREBP pathway protein sequence.

Chimeric or fusion proteins can be made comprising an SREBP pathway protein or fragment thereof (preferably comprising one or more structural or functional domains of the SREBP pathway protein) joined at its amino- or carboxy-terminus via a peptide bond to an amino acid sequence of a different protein. Chimeric proteins can be produced by any known method. These include: recombinant expression of a nucleic acid encoding the protein (comprising an SREBP pathway coding sequence joined in-frame to a coding sequence for a different protein), ligating the appropriate nucleic acid sequences encoding the desired amino acid sequences to each other in the proper coding frame and expressing the chimeric product, and protein synthetic techniques (*e.g.* by use of a peptide synthesizer).

SREBP pathway gene regulatory elements

SREBP pathway gene regulatory DNA elements, such as enhancers or promoters can be used to identify tissues, cells, genes and factors that specifically control SREBP pathway protein production. In the case of dS1P, for instance, such regulatory elements reside within nucleotides 1 to 61 of SEQ ID NO:7. Preferably at least 20, more preferably at least 25, and most preferably at least 50 contiguous nucleotides within nucleotides 1 to 61 of SEQ ID NO:7 are used. Analyzing components that are specific to SREBP pathway protein function can lead to an understanding of how to manipulate these regulatory processes, especially therapeutic applications, as well as an understanding of how to diagnose dysfunction in these processes.

Gene fusions with the SREBP pathway regulatory elements can be made. For compact genes that have relatively few and small intervening sequences, such as those described herein for *Drosophila*, the regulatory elements that control spatial and temporal expression patterns are typically found in the DNA immediately upstream of the coding region, extending to the nearest neighboring gene. Regulatory regions can be used to construct gene fusions where the regulatory DNAs are operably fused to a coding region for a reporter protein whose expression is easily detected, and these constructs are introduced as transgenes into the animal of choice. An entire regulatory DNA region can be used, or the regulatory region can be divided into smaller segments to identify sub-elements that might be specific for controlling expression a given cell type or stage of development. Reporter proteins that can be used for construction of these gene fusions include *E. coli* beta-galactosidase and green fluorescent protein (GFP). These can be detected readily *in situ*, and thus are useful

for histological studies and can be used to sort cells that express SREBP pathway proteins (O'Kane and Gehring PNAS (1987) 84(24):9123-9127; Chalfie *et al.*, Science (1994) 263:802-805; and Cumberledge and Krasnow (1994) Methods in Cell Biology 44:143-159). Recombinase proteins, such as FLP or cre, can be used in controlling gene expression through site-specific recombination (Golic and Lindquist (1989) Cell 59(3):499-509; White *et al.*, Science (1996) 271:805-807). Toxic proteins, such as the *reaper* and *hid* cell death proteins are useful to specifically ablate cells that normally express SREBP pathway proteins in order to assess the physiological function of the cells. (Kingston, In Current Protocols in Molecular Biology (1998) Ausubel *et al.*, John Wiley & Sons, Inc. sections 12.0.3-12.10).

Alternatively, a binary reporter system can be used, similar to that described further below, where an SREBP pathway gene's regulatory element is operably fused to the coding region of an exogenous transcriptional activator protein, such as the GAL4 or tTA activators described below, to create a SREBP pathway "driver gene". For the other half of the binary system the exogenous activator controls a separate "target gene" containing a coding region of a reporter protein operably fused to a cognate regulatory element for the exogenous activator protein, such as UAS_G or a tTA-response element, respectively. An advantage of a binary system is that a single driver gene construct can be used to activate transcription from pre-constructed target genes encoding different reporter proteins, each with its own uses as delineated above.

Reporter gene fusions with an SREBP pathway gene's regulatory element are also useful for testing genetic interactions, in order to identify genes that control the expression of SREBP pathway genes, or promote the growth and differentiation of the tissues that expresses the SREBP pathway protein. SREBP pathway gene regulatory DNA elements are also useful in protein-DNA binding assays to identify gene regulatory proteins that control the expression of SREBP pathway genes. The gene regulatory proteins can be detected using a variety of methods that probe specific protein-DNA interactions well known to those skilled in the art (Kingston, *supra*). These include *in vivo* footprinting assays based on protection of DNA sequences from chemical and enzymatic modification within living or permeabilized cells or protection of DNA sequences from chemical or enzymatic modification using protein extracts, nitrocellulose filter-binding assays, and gel electrophoresis mobility shift assays using radioactively labeled regulatory DNA elements mixed with protein extracts. Candidate SREBP pathway gene regulatory proteins can be purified using a combination of conventional and DNA-affinity purification techniques. Molecular cloning strategies can also be used to identify proteins that specifically bind SREBP pathway gene regulatory DNA elements. For example, a *Drosophila* cDNA library in

an expression vector can be screened for cDNAs that encode SREBP pathway gene regulatory element DNA-binding activity. Similarly, the yeast "one-hybrid" system can be used (Li and Herskowitz, Science (1993) 262:1870-1874; Luo *et al.*, Biotechniques (1996) 20(4):564-568; Vidal *et al.*, PNAS (1996) 93(19):10315-10320).

Identification of molecules that interact with SREBP pathway proteins

A variety of methods can be used to identify or screen for molecules, such as proteins or small molecules, which interact with SREBP pathway proteins, or with derivatives or fragments thereof. The assays may employ a purified SREBP pathway protein, or cell lines or model organisms such as *Drosophila* and *C. elegans* that have been genetically engineered to express an SREBP pathway protein. Suitable screening methodologies are well known in the art to test for proteins and other molecules that interact with SREBP pathway genes and proteins (see *e.g.*, PCT International Publication No. WO 96/34099). The newly identified interacting molecules may provide new targets for pharmaceutical agents. A variety of exogenous molecules, both naturally occurring and/or synthetic (*e.g.*, libraries of small molecules or peptides, or phage display libraries), may be screened for binding capacity. In a typical binding experiment, the SREBP pathway protein or fragment is mixed with candidate molecules under conditions conducive to binding, sufficient time is allowed for any binding to occur, and assays are performed to test for bound complexes. Assays to find interacting proteins can be performed by any method known in the art. Examples include immunoprecipitation with an antibody that binds to the protein in a complex followed by analysis by size fractionation of the immunoprecipitated proteins (*e.g.* by denaturing or nondenaturing polyacrylamide gel electrophoresis), Western analysis, non-denaturing gel electrophoresis, *etc.*

Identification of interacting proteins

Two-hybrid assay systems

A preferred method for identifying interacting proteins is a two-hybrid assay system or variation thereof (Fields and Song, Nature (1989) 340:245-246; U.S. Pat. No. 5,283,173; for review see Brent and Finley, Annu. Rev. Genet. (1997) 31:663-704). The most commonly used two-hybrid screen system is performed using yeast. All systems share three elements: 1) a gene that directs the synthesis of a "bait" protein fused to a DNA binding domain; 2) one or more "reporter" genes having an upstream binding site for the bait, and 3) a gene that directs the synthesis of a "prey" protein fused to an activation domain that activates transcription of the reporter gene.

For screening proteins that interact with an SREBP pathway protein, the "bait" is preferably an SREBP pathway protein, expressed as a fusion protein to a DNA binding domain.

The "prey" protein is a protein to be tested for ability to interact with the bait, which is expressed as a fusion protein to a transcription activation domain. The prey proteins can be obtained from recombinant biological libraries expressing random peptides.

The bait fusion protein can be constructed using any suitable DNA binding domain, such as the *E. coli* LexA repressor protein, or the yeast GAL4 protein (Bartel *et al.*, BioTechniques (1993) 14:920-924, Chasman *et al.*, Mol. Cell. Biol. (1989) 9:4746-4749; Ma *et al.*, Cell (1987) 48:847-853; Ptashne *et al.*, Nature (1990) 346:329-331).

The prey fusion protein can be constructed using any suitable activation domain such as GAL4, VP-16, *etc.* The preys may contain useful moieties such as nuclear localization signals (Ylikomi *et al.*, EMBO J. (1992) 11:3681-3694; Dingwall and Laskey, Trends Biochem. Sci. Trends Biochem. Sci. (1991) 16:479-481) or epitope tags (Allen *et al.*, Trends Biochem. Sci. Trends Biochem. Sci. (1995) 20:511-516) to facilitate isolation of the encoded proteins.

A limitation of the two-hybrid system occurs when transmembrane portions of proteins in the bait or the prey fusions are used. This occurs because most two-hybrid systems are designed to function by formation of a functional transcription activator complex within the nucleus. The use of transmembrane portions of the protein can interfere with proper association, folding, and nuclear transport of bait or prey segments (Ausubel *et al.*, *supra*; Allen *et al.*, Trends Biochem. Sci. (1995) 20:511-516). Since SREBP, SCAP, S1P, and S2P all contain membrane-spanning domains, the "bait" is preferably an SREBP pathway protein derivative or a fragment that lacks transmembrane domains.

Any reporter gene can be used that has a detectable phenotype such as reporter genes that allow cells expressing them to be selected by growth on appropriate medium (*e.g.* HIS3, LEU2 described by Chien *et al.*, PNAS (1991) 88:9572-9582; and Gyuris *et al.*, Cell (1993) 75:791-803). Other reporter genes, such as LacZ and GFP, allow cells expressing them to be visually screened (Chien *et al.*, *supra*).

After interacting proteins have been identified, the DNA sequences encoding the proteins can be isolated.

Antibodies and immunoassays

SREBP pathway proteins encoded by SEQ ID NOs:2, 4 , 6 and 8, and derivatives and fragments thereof, such as those discussed above, may be used as immunogens to generate monoclonal or polyclonal antibodies and antibody fragments or derivatives (*e.g.* chimeric, single chain, Fab fragments). Antibodies to a particular domain of an SREBP pathway protein, such as the SRE binding domain, may be desired. In a specific embodiment, fragments of an SREBP pathway protein identified as hydrophilic are used as immunogens for antibody production using art-known methods. Various known methods for antibody production include production of monoclonal antibodies in germ-free animals (PCT/US90/02545), human or other hybridomas (Cole *et al.*, PNAS (1983) 80:2026-2030; Cole *et al.*, in Monoclonal Antibodies and Cancer Therapy (1985) Alan R. Liss, pp. 77-96), and production of humanized antibodies (Jones *et al.*, Nature (1986) 321:522-525; U.S. Pat. 5,530,101). In a particular embodiment, SREBP pathway polypeptide fragments provide specific antigens and/or immunogens, especially when coupled to carrier proteins. For example, peptides are covalently coupled to keyhole limpet antigen (KLH) and the conjugate is emulsified in Freund's complete adjuvant. Laboratory rabbits are immunized according to conventional protocol and bled. The presence of specific antibodies is assayed by solid phase immunosorbent assays using immobilized corresponding polypeptide. Specific activity or function of the antibodies produced may be determined by convenient *in vitro,* cell-based, or *in vivo* assays: *e.g. in vitro* binding assays, *etc.* Binding affinity may be assayed by determination of equilibrium constants of antigen-antibody association (usually at least about 10^7 M^{-1}, preferably at least about 10^8 M^{-1}, more preferably at least about 10^9 M^{-1}).

Immunoassays can be used to identify proteins that interact with or bind to SREBP pathway proteins. Various assays are available for testing the ability of a protein to bind to or compete with binding to a wild-type SREBP pathway protein or for binding to an anti-SREBP pathway protein antibody. Suitable assays include radioimmunoassays, ELISA (enzyme linked immunosorbent assay), immunoradiometric assays, gel diffusion precipitin reactions, immunodiffusion assays, *in situ* immunoassays (*e.g.*, using colloidal gold, enzyme or radioisotope labels), western blots, precipitation reactions, agglutination assays (*e.g.*, gel agglutination assays, hemagglutination assays), complement fixation assays, immunofluorescence assays, protein A assays, immunoelectrophoresis assays, *etc.*

Identification of interacting compounds

Once new SREBP pathway genes or SREBP pathway interacting genes are identified, they can be assessed as potential drug or pesticide targets. Putative drugs and molecules can be applied onto whole insects, nematodes, and other small invertebrate metazoans, and the ability of the compounds to modulate (*e.g.* block or enhance) SREBP pathway activity can be observed. Alternatively, the effect of various compounds on SREBP pathways can be assayed using cells that have been engineered to express one or more SREBP pathways and associated proteins.

Assays of compounds on nematodes

In a typical worm assay, the compounds to be tested are dissolved in DMSO or other organic solvent, mixed with a bacterial suspension at various test concentrations, preferably OP50 strain of bacteria (Brenner, Genetics (1974) 110:421-440), and supplied as food to the worms. The population of worms to be treated can be synchronized larvae (Sulston and Hodgkin, in The Nematode *C. elegans* (1988) Wood, *ed.*, Cold Spring Harbor Laboratory Press, Cold Spring harbor, New York) or adults or a mixed-stage population of animals.

Adult and larval worms are treated with different concentrations of compounds, typically ranging from 1 mg/ml to 0.001 mg/ml. Behavioral aberrations, such as a decrease in motility and growth, and morphological aberrations, sterility, and death are examined in both acutely and chronically treated adult and larval worms. For the acute assay, larval and adult worms are examined immediately after application of the compound and re-examined periodically (every 30 minutes) for 5-6 hours. Chronic or long-term assays are performed on worms and the behavior of the treated worms is examined every 8-12 hours for 4-5 days. In some circumstances, it is necessary to reapply the compound to the treated worms every 24 hours for maximal effect.

Assays of compounds on insects

Compounds can be administered to insects in a variety of ways, including orally (including addition to synthetic diet, application to plants or prey to be consumed by the test organism), topically (including spraying, direct application of compound to animal, allowing animal to contact a treated surface), or by injection. Hydrophobic molecules must commonly be dissolved in organic solvents, which are allowed to evaporate in the case of methanol or acetone, or at low concentrations can be included to facilitate uptake (ethanol, dimethyl sulfoxide).

The first step in an insect assay is usually the determination of the minimal lethal dose (MLD) on the insects after a chronic exposure to the compounds. The compounds are usually diluted in DMSO, and applied to the food surface bearing 0-48 hour old embryos and larvae. In addition to MLD, this step allows the determination of the fraction of eggs that hatch, behavior of the larvae, such as how they move /feed compared to untreated larvae, the fraction that survive to pupate, and the fraction that eclose (emergence of the adult insect from puparium). Based on these results more detailed assays with shorter exposure times may be designed, and larvae might be dissected to look for obvious morphological defects. Once the MLD is determined, more specific acute and chronic assays can be designed.

In a typical acute assay, compounds are applied to the food surface for embryos, larvae, or adults, and the animals are observed after 2 hours and after an overnight incubation. For application on embryos, defects in development and the percent that survive to adulthood are determined. For larvae, defects in behavior, locomotion, and molting may be observed. For application on adults, behavior and neurological defects are observed, and effects on fertility are noted.

For a chronic exposure assay, adults are placed on vials containing the compounds for 48 hours, then transferred to a clean container and observed for fertility, neurological defects, and death.

Assay of compounds using cell cultures

Compounds that modulate (*e.g.* block or enhance) SREBP pathway activity may also be assayed using cell culture. In one embodiment, compounds that alter the protease activity of S1P are tested. Various compounds added to cells expressing dS1P and/or other SREBP pathway genes may be screened for their ability to modulate the activity of SREBP pathway genes based upon measurements of protease activity. Assays for changes in protease activity can be performed on cultured cells expressing endogenous normal or mutant SREBP pathway genes. Such studies also can be performed on cells transfected with vectors capable of expressing these genes, or their functional domains. In addition, to enhance the signal measured in such assays, cells may be cotransfected with genes encoding dS1P proteins.

For example, cells may be transfected with soluble or membrane bound dS1P, and lysed. The lysates may be analyzed for ability to process peptides corresponding to cleavage of fly SREBP at site 1, in presence or absence of compounds, using a fluorogenic peptide assay essentially as described (Cheng D. *et al.*, J. Biol. Chem. (1999) 274:22805-22812). Compounds that selectively modulate the dS1P activity are identified as potential drug candidates having dS1P specificity.

Identification of small molecules and compounds as potential pharmaceutical compounds from large chemical libraries requires high-throughput screening (HTS) methods (Bolger, Drug Discovery Today (1999) 4:251-253). Several of the assays mentioned herein can lend themselves to such screening methods. For example, cells or cell lines expressing wild type or mutant dS1P protein or derivatives, and a reporter gene, can be treated with compounds of interest. Interactions can be measured using a variety of methods depending on the reporter genes, such as color detection, fluorescence detection (*e.g.* GFP), autoradiography, scintillation analysis, *etc.*

***In vivo* and *in vitro* models of SREBP pathway gene function and dysfunction**

Both *in vivo* models, genetically modified animal models such as *C. elegans* and *Drosophila*, and *in vitro* models such as genetically engineered cell lines expressing or mis-expressing SREBP pathway genes, are useful for studying lipid metabolism and disorders associated with abnormal lipid metabolism. Such models that display detectable phenotypes, such as those described in more detail below and in the examples, can be used for the identification and characterization of SREBP pathway genes or other genes of interest and/or phenotypes associated with the mutation or mis-expression of an SREBP pathway protein. The term "mis-expression" as used herein encompasses mis-expression due to gene mutations. Thus, a mis-expressed SREBP pathway protein may be one having an amino acid sequence that differs from wild type (*i.e.* it is a derivative of the normal protein). A mis-expressed SREBP pathway protein may also be one in which one or more amino acids have been deleted, and thus is a "fragment" of the normal protein. As used herein, "mis-expression" also includes over-expression (*e.g.* by multiple gene copies), underexpression, and non-expression (*e.g.* by gene knockout or blocking expression that would otherwise normally occur). As used in the following discussion concerning *in vivo* and *in vitro* models, the term "gene of interest" refers to an SREBP pathway gene (*i.e.* SREBP, SCAP, S1P, and S2P), or any gene involved in regulation or modulation of the SREBP pathway. Such genes may include any gene involved in the biosynthesis or metabolism of cholesterol or fatty acids such as HMG coenzyme A synthase, HMG-CoA reductase, farnesyl diphosphate synthase, squalene synthase, fatty acid synthase, acetyl-CoA carboxylase, glycerol-3-phosphate acyltransferase, acyl-CoA binding protein, stearoyl CoA desaturase-1, lipoprotein lipase, and the LDL receptor.

The *in vivo* and *in vitro* models may be genetically engineered or modified so that they 1) have deletions and/or insertions of one or more SREBP pathway genes, 2) harbor interfering RNA sequences derived from SREBP pathway genes, 3) have had

one or more endogenous SREBP pathway genes mutated (*e.g.* contain deletions, insertions, rearrangements, or point mutations in SREBP pathway gene or other genes in the pathway), and/or 4) contain transgenes for mis-expression of wild-type or mutant forms of such genes. Such genetically modified *in vivo* and *in vitro* models are useful for identification of new genes that are involved in the synthesis, activation, control, *etc.* of SREBP pathway genes and/or gene products. Further, other genes of interest that are involved in cholesterol and/or fatty acid biosynthesis or metabolism may be identified. The newly identified genes could constitute possible pesticide targets (as judged by animal model phenotypes such as non-viability, block of normal development, defective feeding, defective movement, or defective reproduction). Alternatively, or additionally, they may constitute possible therapeutic targets, particularly in the area of metabolic diseases and disorders, for example, cholesterol synthesis, metabolism, and other fatty acid disorders. The model systems can also be used to test potential pesticidal or pharmaceutical compounds that interact with the SREBP pathway. For example, the compound can be administered to the model system using any suitable method (*e.g.* direct contact, ingestion, injection), and any changes in phenotype, such as changes in lipid content or lethality, can be observed. Various genetic engineering and expression modification methods that can be used are well known in the art, including chemical mutagenesis, transposon mutagenesis, antisense RNAi, dsRNAi, and transgene-mediated mis-expression.

Generating loss-of-function mutations by mutagenesis

Loss-of-function mutations in an invertebrate metazoan SREBP pathway genes can be generated by any of several mutagenesis methods known in the art (Ashburner, *Drosophila melanogaster*: A Laboratory Manual (1989), Cold Spring Harbor, NY, Cold Spring Harbor Laboratory Press: pp. 299-418; Greenspan RJ, Fly pushing: The Theory and Practice of *Drosophila melanogaster* Genetics (1997) Cold Spring Harbor Press, Plainview, NY; Wood *supra*). Techniques for producing mutations in a gene or genome include use of radiation (*e.g.*, X-ray, UV, or gamma ray); chemicals (*e.g.*, EMS, MMS, ENU, formaldehyde, *etc.*); and insertional mutagenesis by mobile elements including dysgenesis induced by transposon insertions, or transposon-mediated deletions, for example, male recombination, as described below. Other methods of reducing expression of genes include antisense; double-stranded RNA interference; peptide and RNA aptamers; directed deletions; homologous recombination; dominant negative alleles; and intrabodies.

Generating loss-function phenotypes by transposon insertion or excision

Transposable elements are particularly useful for inserting sequences into a gene of interest so that the encoded protein is not properly expressed, creating a "knock-out" animal having a loss-of-function phenotype. There are several suitable transposable elements that can be used. Techniques are well-established for the use of P element in *Drosophila* (Rubin and Spradling, Science (1982) 218:348-53; U.S. Pat. No. 4,670,388) and Tc1 in *C. elegans* (Zwaal *et al.*, Proc. Natl. Acad. Sci. U.S.A. (1993) 90:7431-7435; and *Caenorhabditis elegans*: Modern Biological Analysis of an Organism (1995) Epstein and Shakes, *eds.*). Other Tc1-like transposable elements can be used such as minos, mariner and sleeping beauty, as well transposable elements that function in a variety of insect species, such as *PiggyBac*, *hobo*, and *hermes* (Thibault *et al.*, Insect Mol Biol (1999) 8:119-23).

P elements, which contain one or more elements that allow detection of animals containing the P element, are preferred for the isolation of loss-of-function mutations in *Drosophila* SREBP pathway genes. Most often, marker genes are used that affect the eye color of *Drosophila*, such as derivatives of the *Drosophila white* or *rosy* genes (Rubin and Spradling, Science (1982) 218(4570):348-353; and Klemenz *et al.*, Nucleic Acids Res. (1987) 15(10):3947-3959). However, in principle, any gene can be used as a marker that causes a reliable and easily scored phenotypic change in transgenic animals. Various other markers include bacterial plasmid sequences having selectable markers such as ampicillin resistance (Steller and Pirrotta, EMBO. J. (1985) 4:167-171); and *lacZ* sequences fused to a weak general promoter to detect the presence of enhancers with a developmental expression pattern of interest (Bellen *et al.*, Genes Dev. (1989) 3(9):1288-1300). Other examples of marked P elements useful for mutagenesis have been reported (Nucleic Acids Research (1998) 26:85-88; and http://flybase.bio.indiana.edu).

Recently, a transgenesis system was described that may be suitable for following gene transfer in eye-bearing animals of diverse insect and other species (Berghammer *et al.*, Nature (1999) 402:370-371). An artificial promoter that is universally expressed and restricted to eye tissue drives expression of a universal marker, such as GFP. The preferred promoter contains three binding sites for Pax-6 homodimers in front of a TATA box, (Sheng *et al.*, Genes Devel. (1997) 11:1122-1131). The system is transferred via a vector capable of delivering transgenes to a broad range of animal species, such as transposon-based vectors derived from *Hermes*, *piggyBac*, or *mariner*, or pantropic retroviruses (Jordan *et al.*, Insect Mol Biol (1998) 7: 215-222; U.S. Pat. No.5,670,345).

Various research groups have created large collections of P-insertions that generate new loss-of-function mutations upon random insertion in the genome (Spradling *et al.*, Genetics (1999) 153:135-77). Since the sequence of the P elements is known, the genomic sequence flanking each transposon insert is determined either by plasmid rescue (Hamilton *et al.*, PNAS (1991) 88:2731-2735) or by inverse polymerase chain reaction (Rehm, http://www.fruitfly.org/methods/) to determine if the P is in a gene of interest.

A preferred method of transposon mutagenesis in *Drosophila* employs the "local hopping" method described by Tower *et al.* (Genetics (1993) 133:347-359). Each new P insertion line can be tested molecularly for transposition of the P element into the gene of interest (*e.g.* dS1P) by assays based on PCR. For each reaction, one PCR primer is used that is homologous to sequences contained within the P element and a second primer is homologous to the coding region or flanking regions of the gene of interest. Products of the PCR reactions are detected by agarose gel electrophoresis. The sizes of the resulting DNA fragments reveal the site of P element insertion relative to the gene of interest. Alternatively, Southern blotting and restriction mapping using DNA probes derived from genomic DNA or cDNAs of the gene of interest can be used to detect transposition events that rearrange the genomic DNA of the gene. P transposition events that map to the gene of interest can be assessed for phenotypic effects in heterozygous or homozygous mutant *Drosophila*.

In another embodiment, *Drosophila* lines carrying P insertions in the gene of interest, can be used to generate localized deletions using "imprecise excision" methods (Kaiser, Bioessays (1990) 12(6):297-301; Harnessing the power of *Drosophila* genetics, In Drosophila melanogaster: Practical Uses in Cell and Molecular Biology, Goldstein and Fyrberg, Eds., Academic Press, Inc. San Diego, California). This is particularly useful if no P element transpositions are found that disrupt the gene of interest. Briefly, flies containing P elements inserted near the gene of interest are exposed to a further round of transposase to induce excision of the element. Progeny in which the transposon has excised are typically identified by loss of the eye color marker associated with the transposable element. The resulting progeny will include flies with either precise or imprecise excision of the P element, where the imprecise excision events often result in deletion of genomic DNA neighboring the site of P insertion. Such progeny are screened by molecular techniques to identify deletion events that remove genomic sequence from the gene of interest, and assessed for phenotypic effects in heterozygous and homozygous mutant *Drosophila*. We have used this method to generate a small deletion in the enhancer/5' UTR region of *Drosophila* SREBP. While the original P-element insertion in this

region was not lethal to the animal, the corresponding excision, which left most of the P-element intact and removed ~450 bp of genomic DNA, was homozygous lethal and was lethal over a large deficiency that removes *Drosophila* SREBP. The deletion allele may be used to isolate further mutations in the gene and to further investigate the function of *Drosophila* SREBP.

P-element mobilization can further generate chromosomal deletions and duplications by the process of "male recombination (Preston *et al.*, Genetics (1996) 144:1623-38)". P-mediated germ line recombination (*i.e.*, cross-over events between paired chromosomes in the germline) occurs in both males and females but is more conspicuous in the males, where recombination is usually absent. Male recombination frequently induces deletions and reciprocal duplications of the adjacent chromosomal DNA.

In *C. elegans*, Tc1 transposable element can be used for directed mutagenesis of a gene of interest. Typically, a Tc1 library is prepared by the methods of Zwaal *et al.*, *supra* and Plasterk, *supra*, using a strain in which the Tc1 transposable element is highly mobile and present in a high copy number. The library is screened for Tc1 insertions in the region of interest using PCR with one set of primers specific for Tc1 sequence and one set of gene-specific primers and *C. elegans* strains that contain Tc1 transposon insertions within the gene of interest are isolated. Frequently, Tc1 insertions do not fully disrupt the function of a gene due to insertion into non-coding sequence or the ability of the host transcriptional machinery to bypass the effect of the insertion. In this case, imprecise excision of the Tc1 element may be used to identify a deletion in the gene of interest. As described in detail in the Examples, we used this method of Tc1 insertion followed by imprecise excision to generate a partial deletion allele of ceSREBP.

Generating loss-of-function phenotypes using RNA-based methods

SREBP pathway genes may be identified and/or characterized by generating loss-of-function phenotypes in animals of interest through RNA-based methods, such as antisense RNA (Schubiger and Edgar, Methods in Cell Biology (1994) 44:697-713). One form of the antisense RNA method involves the injection of embryos with an antisense RNA that is partially homologous to the gene of interest (in this case any of the SREBP pathways genes of SEQ ID NO: 1, 3, 5, or 7). Another form of the antisense RNA method involves expression of an antisense RNA partially homologous to the gene of interest by operably joining a portion of the gene of interest in the antisense orientation to a powerful promoter that can drive the expression of large quantities of antisense RNA, either generally throughout the animal or in

specific tissues. Antisense RNA-generated loss-of-function phenotypes have been reported previously for several *Drosophila* genes including *cactus*, *pecanex*, and *Krüppel* (LaBonne *et al.*, Dev. Biol. (1989) 136(1):1-16; Schuh and Jackle, Genome (1989) 31(1):422-425; Geisler *et al.*, Cell (1992) 71(4):613-621).

Loss-of-function phenotypes can also be generated by cosuppression methods (Bingham Cell (1997) 90(3):385-387; Smyth, Curr. Biol. (1997) 7(12):793-795; Que and Jorgensen, Dev. Genet. (1998) 22(1):100-109). Cosuppression is a phenomenon of reduced gene expression produced by expression or injection of a sense strand RNA corresponding to a partial segment of the gene of interest. Cosuppression effects have been employed extensively in plants and *C. elegans* to generate loss-of-function phenotypes, and there is a single report of cosuppression in *Drosophila*, where reduced expression of the Adh gene was induced from a *white-Adh* transgene using cosuppression methods (Pal-Bhadra *et al.*, Cell (1997) 90(3):479-490).

Another method for generating loss-of-function phenotypes is by double-stranded RNA interference. This method is based on the interfering properties of double-stranded RNA derived from the coding regions of gene, and has proven to be of great utility in genetic studies of *C. elegans* (Fire *et al.*, Nature (1998) 391:806-811), and can also be used to generate loss-of-function phenotypes in *Drosophila* (Kennerdell and Carthew, Cell (1998) 95:1017-1026; Misquitta and Patterson PNAS (1999) 96:1451-1456). In one example of this method, complementary sense and antisense RNAs derived from a substantial portion of a gene of interest, such as an SREBP pathway gene, are synthesized *in vitro*. The resulting sense and antisense RNAs are annealed in an injection buffer, and the double-stranded RNA injected or otherwise introduced into animals. Progeny of the injected animals are then inspected for phenotypes of interest (PCT publication no. WO99/32619). In another embodiment of the method, the dsRNA can be delivered to the animal by bathing the animal in a solution containing a sufficient concentration of the dsRNA. In another embodiment of the method, dsRNA derived from SREBP pathway genes can be generated *in vivo* by simultaneous expression of both sense and antisense RNA from appropriately positioned promoters operably fused to the SREBP pathway gene sequence in both sense and antisense orientations. In yet another embodiment of the method the dsRNA can be delivered to the animal by engineering expression of dsRNA within cells of a second organism that serves as food for the animal. Examples include engineering expression of dsRNA in *E. coli* bacteria that are fed to *C. elegans*, engineering expression of dsRNA in baker's yeast which is fed to *Drosophila*, or engineering expression of dsRNA in transgenic plants that are fed to plant-eating insects such as *Leptinotarsa* or *Heliothis*.

We have used dsRNAi injection, soaking, and feeding methods in *C. elegans* to generate specific intestinal phenotypes, as discussed in detail in Example 5. We have also used dsRNAi injection into *Drosophila* embryos. For this experiment, dsRNA fragments were generated that spanned the amino-terminal ~2kb of the coding sequence of *Drosophila* SREBP. Animals that were injected as early embryos generally died during larval development, demonstrating that *Drosophila* SREBP is essential for viability.

Recently, RNAi has been successfully used in cultured *Drosophila* cells to inhibit expression of targeted proteins (Dixon lab, University of Michigan, http://dixonlab.biochem.med.umich.edu/protocols/RNAiExperiments.html). Thus, cell lines in culture can be manipulated using RNAi both to perturb and study the function of SREBP pathway components and to validate the efficacy of therapeutic strategies that involve the manipulation of this pathway.

The observation that RNAi can selectively inactivate one or more genes simultaneously in *Drosophila* tissue culture cells provides an additional means of screening for genes that regulate SREBP activation. In this screen RNAi suppresses expression of a regulatory gene and thus mimics the function of a drug that specifically targets that gene function. The procedure for carrying out RNAi in *Drosophila* tissue culture cells can be carried out on a large scale such that part or all the genes in the *Drosophila* genome are inactivated by RNAi and tested as potential regulatory genes. Based on ESTs and sequence from the complete *Drosophila* genome, a library of double stranded RNAi molecules directed against these genes is constructed by *in vitro* transcription from convergently oriented T7 promoters. The sense and antisense RNA strands are annealed and arrayed into 96-well plates with each well containing a different RNAi molecule directed against a different gene in the genome.

Aliquots of *Drosophila* tissue culture cells, such as wild type S2 or S2 cells specifically engineered to assay for SREBP function, are placed in the wells of a 96-well plate. An aliquot of each of the 96 different dsRNAs from the RNAi library is added to a different well containing the *Drosophila* cells. Each well of the 96 well plate thus contains a different RNAi molecule directed against a different gene in the genome. Following the introduction of the RNAi library, the cells are assayed for SREBP activation. Such assays can measure a decrease in membrane-bound SREBP, SREBP translocation to the nucleus, SREBP phosphorylation, or increased transcription of SREBP transcriptional or indirect target genes (such as fatty acid synthase or ATP citrate lyase – as described in the section "Assays for Change in Gene Expression"). A change in the activation of SREBP in RNAi treated cells and

not control cells indicates that the RNAi molecule directed against that regulator gene is causing this change. Assays can be designed to uncover genes that either up-regulate or down-regulate the activation of SREBP.

Generating loss-of-function phenotypes using peptide and RNA aptamers

Another method for generating loss-of-function phenotypes is by the use of peptide aptamers, which are peptides or small polypeptides that act as dominant inhibitors of protein function. Peptide aptamers specifically bind to target proteins, blocking their function ability (Kolonin and Finley, PNAS (1998) 95:14266-14271). Due to the highly selective nature of peptide aptamers, they may be used not only to target a specific protein, but also to target specific functions of a given protein (*e.g.* proteolytic function). Further, peptide aptamers may be expressed in a controlled fashion by use of promoters which regulate expression in a temporal, spatial or inducible manner. Peptide aptamers act dominantly; therefore, they can be used to analyze proteins for which loss-of-function mutants are not available.

Peptide aptamers that bind with high affinity and specificity to a target protein may be isolated by a variety of techniques known in the art. In one method, they are isolated from random peptide libraries by yeast two-hybrid screens (Xu *et al.*, PNAS (1997) 94:12473-12478). They can also be isolated from phage libraries (Hoogenboom *et al.*, Immunotechnology (1998) 4:1-20) or chemically generated peptides/libraries.

RNA aptamers are specific RNA ligands for proteins that can specifically inhibit protein function of the gene (Good *et al.*, Gene Therapy (1997) 4:45-54; Ellington. *et al.*, Biotechnol. Annu. Rev. (1995) 1:185-214). *In vitro* selection methods can be used to identify RNA aptamers having a selected specificity (Bell *et al.*, J. Biol. Chem. (1998) 273:14309-14314). It has been demonstrated that RNA aptamers can inhibit protein function in *Drosophila* (Shi *et al.*, Proc. Natl. Acad. Sci USA (1999) 96:10033-10038). Accordingly, RNA aptamers can be used to decrease the expression of an SREBP pathway protein or derivative thereof, or a protein that interacts with an SREBP pathway protein.

Transgenic animals can be generated to test peptide or RNA aptamers *in vivo* (Kolonin, MG, and Finley, RL, Genetics, (1998) 95:4266-4271). For example, transgenic *Drosophila* lines expressing the desired aptamers may be generated by P element mediated transformation (discussed below). The phenotypes of the progeny expressing the aptamers can then be characterized.

Generating loss of function phenotypes using intrabodies

Intracellularly expressed antibodies, or intrabodies, are single-chain antibody molecules designed to specifically bind and inactivate target molecules inside cells. Intrabodies have been used in cell assays and in whole organisms such as *Drosophila* (Chen *et al.*, Hum. Gen. Ther. (1994) 5:595-601; Hassanzadeh *et al.*, Febs Lett. (1998) 16(1, 2):75-80 and 81-86). Inducible expression vectors can be constructed with intrabodies that react specifically with an SREBP pathway protein. These vectors can be introduced into model organisms and studied in the same manner as described above for aptamers.

Transgenesis

Typically, transgenic animals are created that contain gene fusions of the coding regions of an SREBP pathway gene (from either genomic DNA or cDNA) or genes engineered to encode antisense RNAs, cosuppression RNAs, interfering dsRNA, RNA aptamers, peptide aptamers, or intrabodies operably joined to a specific promoter and transcriptional enhancer whose regulation has been well characterized, usually heterologous promoters/enhancers (*i.e.* promoters/enhancers that are non-native to the SREBP pathway genes being expressed). Methods are well known for incorporating exogenous nucleic acid sequences into the genome of animals or cultured cells to create transgenic animals or recombinant cell lines. For invertebrate animal models, the most common methods involve the use of transposable elements.

In addition to creating loss-of-function phenotypes, transposable elements can be used to incorporate the gene of interest, or mutant or derivative thereof, as an additional gene into any region of an animal's genome resulting in mis-expression of the gene. A preferred vector designed specifically for misexpression of genes in transgenic *Drosophila*, is derived from pGMR (Hay *et al.*, Development (1994) 120:2121-2129). It is 9Kb long, and contains: an origin of replication for *E. coli*; an ampicillin resistance gene; P element transposon ends to mobilize the inserted sequences; a *White* marker gene; an expression unit comprising the TATA region of hsp70 enhancer and the 3'untranslated region of α-tubulin gene. The expression unit contains a first multiple cloning site (MCS) designed for insertion of an enhancer and a second MCS located 500 bases downstream, designed for the insertion of a gene of interest. Homologous recombination or gene targeting techniques can be used to substitute a gene of interest for one or both copies of the animal's homologous gene. The transgene can be under the regulation of either an exogenous or an endogenous promoter element, and be inserted as either a minigene or a large genomic fragment. In one application, gene function can be analyzed by ectopic expression, using, for

example, *Drosophila* (Brand *et al.*, Methods in Cell Biology (1994) 44:635- 654) or *C. elegans* (Mello and Fire, Methods in Cell Biology (1995) 48:451-482).

Genes are typically introduced into *C. elegans* via injection into the ovaries of young adult hermaphrodites. Circular DNA that is injected into nematodes at high concentrations generally concatamerizes to form "extrachromosomal arrays," which typically contain hundreds of copies of the injected sequences and are stably transmitted at rates of ~10% to 90%. X-ray irradiation of animals carrying the array can induce chromosomal integration of the transgenes (Epstein and Shakes, *supra*). In a preferred embodiment, gene fusions for directing the mis-expression of SREBP pathway genes are incorporated into a vector which is injected into nematodes along with a plasmid containing a dominant selectable marker, such as *rol-6*. Transgenic animals are identified as those exhibiting a roller phenotype, and the transgenic animals are inspected for additional phenotypes of interest created by mis-expression of the SREBP pathway gene.

Examples of well-characterized heterologous promoters that may be used to create the transgenic animals include heat shock promoters/enhancers, which are useful for temperature induced mis-expression. In *Drosophila*, these include the *hsp70* and *hsp83* genes, and in *C. elegans*, include *hsp 16-2* and *hsp 16-41*. Tissue specific promoters/enhancers are also useful, and in *Drosophila*, include *eyeless* (Mozer and Benzer, Development (1994) 120:1049-1058), *sevenless* (Bowtell *et al.*, PNAS (1991) 88(15):6853-6857), and *glass*-responsive promoters/enhancers (Quiring *et al.*, Science (1994) 265:785-789) which are useful for expression in the eye; and enhancers/promoters derived from the *dpp* or *vestigal* genes which are useful for expression in the wing (Staehling-Hampton *et al.*, Cell Growth Differ. (1994) 5(6):585-593; Kim *et al.*, Nature (1996) 382:133-138). In *C. elegans*, examples of useful tissue specific promoters/enhancers include the *myo-2* gene promoter, useful for pharyngeal muscle-specific expression; the *hlh-1* gene promoter, useful for body-muscle-specific expression; and the gene promoter, useful for touch-neuron-specific gene expression. Finally, where it is necessary to restrict the activity of dominant active or dominant negative transgenes to regions where the pathway is normally active, it may be useful to use endogenous promoters of genes in the pathway, such as the SREBP pathway genes. In *Drosophila*, binary control systems that employ exogenous DNA are useful when testing the mis-expression of genes in a wide variety of developmental stage-specific and tissue-specific patterns. Two examples of binary exogenous regulatory systems include the UAS/GAL4 system from yeast (Hay *et al.*, PNAS (1997) 94(10):5195-5200; Ellis *et al.*, Development (1993) 119(3):855-865), and the "Tet system" derived from *E. coli* (Bello et al., Development (1998)

125:2193-2202). The UAS/GAL4 system is a well-established and powerful method of mis-expression in *Drosophila* which employs the UAS_G upstream regulatory sequence for control of promoters by the yeast GAL4 transcriptional activator protein (Brand and Perrimon, Development (1993) 118(2):401-15). In this approach, transgenic *Drosophila*, termed "target" lines, are generated where the gene of interest to be mis-expressed is operably fused to an appropriate promoter controlled by UAS_G. Other transgenic *Drosophila* strains, termed "driver" lines, are generated where the GAL4 coding region is operably fused to promoters/enhancers that direct the expression of the GAL4 activator protein in specific tissues, such as the eye, wing, nervous system, gut, or musculature. The gene of interest is not expressed in the target lines for lack of a transcriptional activator to drive transcription from the promoter joined to the gene of interest. However, when the UAS-target line is crossed with a GAL4 driver line, mis-expression of the gene of interest is induced in resulting progeny in a specific pattern that is characteristic for that GAL4 line. The technical simplicity of this approach makes it possible to sample the effects of directed mis-expression of the gene of interest in a wide variety of tissues by generating one transgenic target line with the gene of interest, and crossing that target line with a panel of pre-existing driver lines.

In the "Tet" binary control system, transgenic *Drosophila* driver lines are generated where the coding region for a tetracycline-controlled transcriptional activator (tTA) is operably fused to promoters/enhancers that direct the expression of tTA in a tissue-specific and/or developmental stage-specific manner. The driver lines are crossed with transgenic *Drosophila* target lines where the coding region for the gene of interest to be mis-expressed is operably fused to a promoter that possesses a tTA-responsive regulatory element. When the resulting progeny are supplied with food supplemented with a sufficient amount of tetracycline, expression of the gene of interest is blocked. Expression of the gene of interest can be induced at will simply by removal of tetracycline from the food. Also, the level of expression of the gene of interest can be adjusted by varying the level of tetracycline in the food. Thus, the use of the Tet system as a binary control mechanism for mis-expression has the advantage of providing a means to control the amplitude and timing of mis-expression of the gene of interest, in addition to spatial control. Consequently, if a gene of interest (*e.g.* an SREBP pathway gene) has lethal or deleterious effects when mis-expressed at an early stage in development, such as the embryonic or larval stages, the function of the gene of interest in the adult can still be assessed by adding tetracycline to the food during early stages of development and removing tetracycline later so as to induce mis-expression only at the adult stage.

Dominant negative mutations, which cause the mutant protein to interfere with the normal function of a wild-type copy of the protein and which can result in loss-of-function or reduced-function phenotypes in the presence of a normal copy of the gene, can be made using known methods (Hershkowitz, Nature (1987) 329:219-222). In the case of active monomeric proteins, overexpression of an inactive form, achieved, for example, by linking the mutant gene to a highly active promoter, can cause competition for natural substrates or ligands sufficient to significantly reduce net activity of the normal protein. Alternatively, changes to active site residues can be made to create a virtually irreversible association with a target.

Assays for change in gene expression

Various expression analysis techniques may be used to identify genes that are differentially expressed between a cell line or an animal expressing a wild type SREBP pathway gene compared to another cell line or animal expressing a mutant SREBP pathway gene. Such expression profiling techniques include differential display, serial analysis of gene expression (SAGE), transcript profiling coupled to a gene database query, nucleic acid array technology, subtractive hybridization, and proteome analysis (*e.g.* mass-spectrometry and two-dimensional protein gels). Nucleic acid array technology may be used to determine a global (i.e., genome-wide) gene expression pattern in a normal animal for comparison with an animal having a mutation in one or more SREBP gene. Gene expression profiling can also be used to identify other genes (or proteins) that may have a functional relation to SREBP pathway genes (*e.g.* may participate in a signaling pathway with these genes). The genes are identified by detecting changes in their expression levels following mutation, *(i.e.*, insertion, deletion or substitution in, or over-expression, under-expression, mis-expression or knock-out) of the SREBP pathway gene.

We used transcriptional profiling to identify genes that are up-regulated in flies that mis-express an activated form of *Drosophila* SREBP (dSREBP.CA) and are thus likely transcriptional targets or indirect targets of SREBP activation in the wild-type animal. As described below in Example 7, we generated a fly that mis-expressed dSREBP.CA in the larval fat body and appeared to have metabolic phenotypes. We isolated mRNA from the mutant and wild-type larvae, generated radioactively-labeled cDNA, and separately probed identical nylon filters that had been spotted with PCR products representing the Drosophila homologues of ~80 mammalian metabolic genes. The signals from all spots on the filters were quantitated on a phosphoimager, and the normalized signals from corresponding spots on filters probed with wild type or mutant mRNA were compared. Our methods were adapted from the Atlas Array

system (Clonetech, Palo Alto, CA). This analysis identified several genes that appeared to be up regulated in the mutant animals. Northern blot analysis, using mRNA from wild-type and mutant larvae and DNA probes made from the candidate target genes, confirmed results of the profiling experiments. The *Drosophila* homologues of fatty acid synthase and ATP citrate lyase were the most strongly up regulated, by factors of 10 and 4, respectively. Both genes contribute to the biosynthesis of fatty acids and are transcriptional targets of mammalian SREBPs.

Knowledge of these targets provides means of assaying *Drosophila* SREBP activity *in vivo* or i*n vitro*. To this end, reporter constructs can be generated that fuse these genes' regulatory elements to a reporter gene, such as GFP or lacZ. Such constructs can be introduced into animals or cells and used to assay genetic or chemical modifications that alter the activity of SREBP pathway proteins.

Phenotypes associated with SREBP pathway gene mutations

After isolati n of model animals carrying mutated or mis-expressed SREBP pathway genes or inhibitory RNAs, animals are carefully examined for phenotypes of interest. For analysis of SREBP pathway genes that have been mutated (i.e. deletions, insertions, and/or point mutations) animal models that are both homozygous and heterozygous for the altered SREBP pathway gene are analyzed. Examples of specific phenotypes that may be investigated include lethality; sterility; and changes in various characteristics of the animal such as motility, body shape, body size and weight, metabolism, lipid accumulation, feeding, development, morphogenesis of organs, brood size, thermotaxis, etc. Some phenotypes more specific to *Drosophila* include alterations in: morphogenesis of the peripheral sensory organs, imaginal discs, eye development, wing development, leg development, bristle development, antennae development, gut development, fat body, and musculature. Some phenotypes more specific to nematodes include: alterations in chemotaxis, a dauer constitutive phenotype, a dauer defective phenotype, and a pale-intestine phenotype. A phenotype of particular interest in *C. elegans* is the pale intestine phenotype, which is indicative of defects in lipid metabolism and is discussed in more detail below and in the Examples. Both the partial deletion allele of ceSREBP, described above, and dsRNAi using ceSREBP gene fragments produces this pale intestine phenotype.

Genomic sequences containing an SREBP pathway gene can be used to confirm whether an existing mutant insect or worm line corresponds to a mutation in one or more SREBP pathway genes, by rescuing the mutant phenotype. Briefly, a genomic fragment containing the SREBP pathway gene of interest and potential flanking regulatory regions can be subcloned into any appropriate insect (such as

Drosophila) or worm (such as *C. elegans*) transformation vector, and injected into the animals. For *Drosophila*, an appropriate helper plasmid is used in the injections to supply transposase for transposon-based vectors. Resulting germline transformants are crossed for complementation testing to an existing or newly created panel of *Drosophila* or *C. elegans* lines whose mutations have been mapped to the vicinity of the gene of interest (Greenspan, *supra*; and *Caenorhabditis elegans*: Modern Biological Analysis of an Organism (1995), Epstein and Shakes, eds.). If the genomic fragment rescues a mutant line, as judged by complementation of the mutant phenotype, then the mutant line likely harbors a mutation in the SREBP pathway gene. This prediction can be further confirmed by sequencing the SREBP pathway gene from the mutant line to identify the lesion in the SREBP pathway gene.

Using above described methods of EMS mutagenesis, phenotypic analysis, and mutation detection, we identified a point mutation in the ceSREBP gene that weakly confers the loss-of-function pale intestine phenotype. Briefly, we mutagenized heterozygous nematodes that carried a partial deletion of ceSREBP, described in detail in Example 4, and, in *trans*, a recessive "dumpy (dp)" gene that conferred a short body. From the progeny, we screened for the rare individuals that displayed the pale intestine phenotype but not the dumpy phenotype, suggesting that the dp chromosome now also contained a mutation in the ceSREBP gene, which, in combination with the deletion allele, produced the intestinal defect. Sequence analysis of genomic DNA from the mutant line confirmed the existence of a point mutation that truncates the carboxy-terminal 36 amino acids of the ceSREBP protein.

Identification of genes that modify SREBP pathway genes

The characterization of new phenotypes created by mutations or misexpression in SREBP pathway genes enables one to test for genetic interactions between SREBP pathway genes and other genes that may participate in the same, related, or interacting genetic or biochemical pathway(s). Individual genes can be used as starting points in large-scale genetic modifier screens as described in more detail below. Alternatively, RNAi methods can be used to simulate loss-of-function mutations in the genes being analyzed. It is of particular interest to investigate whether there are any interactions of SREBP pathway genes with other well-characterized genes, particularly genes involved in lipid metabolism. For example, a candidate gene that may be tested for interaction with the SREBP pathway is the insulin receptor gene (referred to as inr in *Drosophila*, and daf-2 in *C. elegans*).

Genetic modifier screens

A genetic modifier screen using invertebrate model organisms is a particularly preferred method for identifying genes that interact with SREBP pathway genes, because large numbers of animals can be systematically screened making it more likely that interacting genes will be identified. In *C. elegans* and *Drosophila*, a screen of up to about 10,000 animals is considered to be a pilot-scale screen. Moderate-scale screens usually employ about 10,000 to about 50,000 flies or up to about 100,000 worms, and large-scale screens employ greater than about 50,000 or 100,000 flies or worms, respectively. In a genetic modifier screen, animals having a mutant phenotype due to a mutation in one or more SREBP pathway genes are further mutagenized, for example by chemical mutagenesis or transposon mutagenesis. The mutagenesis procedures used in typical genetic modifier screens of *C. elegans* are well known in the art. One method involves exposure of hermaphrodites that carry mutations in one or more SREBP pathway genes to a mutagen, such as EMS or trimethylpsoralen with ultraviolet radiation (Huang and Sternberg, Methods in Cell Biology (1995) 48:97-122). Alternatively, transposable elements are used, oftentimes by the introduction of a mutator locus, such as mut-2, which promotes mobility of transposons (Anderson, Methods in Cell Biology (1995) 4:31-58).

In *Drosophila*, the mutagenesis methods and other procedures used in a genetic modifier screen depend upon the precise nature of the mutant allele being modified; these methods are discussed in more detail below under the *Drosophila* genetic modifier screen subheading.

Progeny of the mutagenized animals are generated and screened for the rare individuals that display suppressed or enhanced versions of the original mutant SREBP pathway phenotype. Such animals are presumed to have mutations in other genes, called "modifier" genes, that participate in the same phenotype-generating pathway. The newly identified modifier genes can be isolated away from the mutations in the SREBP pathway genes by genetic crosses, so that the intrinsic phenotypes caused by the modifier mutations can be assessed in isolation.

We have performed enhancer and suppressor screens in C. elegans to identify genes that interact with the SREBP pathway. For both, the strategy was to mutagenize animals that had a pale intestine phenotype due to defective SREBP gene activity, and to screen for the rare individuals in which the phenotype was altered. For the enhancer screen, we mutagenized animals that were homozygous for the point mutation that truncated the SREBP protein, a weak allele, and screened for progeny that displayed more severe loss of function phenotypes, including intestinal defects, larval arrest, and small body size. In order to identify suppressors, we mutagenized

wild-type animals that were reared on *E. coli* that expressed double-stranded RNA for part of the *C. elegans* SREBP gene. These nematodes developed into fertile adults that displayed pale intestines, reduced body size, and reduced fecundity. In the next generation, we screened for the rare individuals that displayed normal size and intestinal pigmentation. Several modifier genes have been identified.

Modifier genes can be mapped using a combination of genetic and molecular methods known in the art. Modifiers that come from a genetic screen in *C. elegans* are preferably mapped with visible genetic markers and/or with molecular markers such as STS markers (Woods, *supra*; Epstein and Shakes, *supra*). Modifier genes may be uncovered by identification of a genomic clone that rescues the mutant phenotype, as described above. Alternatively, modifier genes that are identified by a Tc1-based screen can be uncovered using transposon display technology (Korswagen et al., Proc Natl Acad Sci U.S.A. (1996) 93(25):14680-5).

Standard techniques used for the mapping of modifiers that come from a genetic screen in *Drosophila* include meiotic mapping with visible or molecular genetic markers, male recombination relative to P-element insertions, complementation analysis with deficiencies, duplications, and lethal P-element insertions, and cytological analysis of chromosomal aberrations (Greenspan, *supra*; Drosophila: A Laboratory Handbook, *supra*). Genes corresponding to modifier mutations that fail to complement a lethal P-element may be cloned by plasmid rescue of the genomic sequence surrounding that P-element. Alternatively, modifier genes may be mapped by phenotype rescue and positional cloning (Sambrook et al., supra).

Newly identified modifier mutations can be tested directly for interaction with other genes of interest known to be involved or implicated in the SREBP using methods described above. Also, the new modifier mutations can be tested for interactions with genes in other pathways that are not believed to be related to lipid metabolism (e.g. *Notch* in *Drosophila*, and *lin* in *C. elegans*). New modifier mutations that exhibit specific genetic interactions with other genes implicated in lipid metabolism, but not interactions with genes in unrelated pathways, are of particular interest.

The modifier mutations may also be used to identify "complementation groups". Two modifier mutations are considered to fall within the same complementation group if animals carrying both mutations in trans exhibit essentially the same phenotype as animals that are homozygous for each mutation individually and, generally, are lethal when in trans to each other. Generally, individual complementation groups defined in this way correspond to individual genes.

When SREBP pathway modifier genes are identified, homologous genes in other species can be isolated using procedures based on cross-hybridization with modifier gene DNA probes, PCR-based strategies with primer sequences derived from the modifier genes, and/or computer searches of sequence databases. For therapeutic applications related to the function of SREBP pathway, human and rodent homologues of the modifier genes are of particular interest. For pesticide and other agricultural applications, homologues of modifier genes in insects and arachnids are of particular interest. Insects, acarids, arachnids, nematodes, and other organisms of interest include, among others, Isopoda; Diplopoda; Chilopoda; Symphyla; Thysanura; Collembola; Orthoptera, such as *Scistocerca* spp; Blattoidea, such as *Blattella germanica*; Dermaptera; Isoptera; Anoplura; Mallophaga; Thysanoptera; Heteroptera; Homoptera, including *Bemisia tabaci*, and *Myzus* spp.; Lepidoptera including *Plodia interpunctella*, *Pectinophora gossypiella*, *Plutella* spp., *Heliothis* spp., and *Spodoptera* species; Coleoptera such as *Leptinotarsa*, *Diabrotica* spp., *Anthonomus* spp., and *Tribolium* spp.; Hymenoptera, including Diptera, including *Anopheles* spp.; Siphonaptera, including *Ctenocephalides felis*; Arachnida; and Acarinan, including *Amblyoma americanum*; and nematodes, including *Meloidogyne* spp., and *Heterodera glycinii*.

Genetic modifier screens in Drosophila

The procedures involved in typical *Drosophila* genetic modifier screens are well known in the art (Wolfner and Goldberg, Methods in Cell Biology (1994) 44:33-80; and Karim *et al.*, Genetics (1996) 143:315-329). The procedures used differ depending upon the precise nature of the mutant allele being modified. If the mutant allele is genetically recessive, as is commonly the situation for a loss-of-function allele, then most typically males, or in some cases females, which carry one copy of the mutant allele are exposed to an effective mutagen, such as EMS, MMS, ENU, triethylamine, diepoxyalkanes, ICR-170, formaldehyde, X-rays, gamma rays, or ultraviolet radiation. The mutagenized animals are crossed to animals of the opposite sex that also carry the mutant allele to be modified. In the case where the mutant allele being modified is genetically dominant, as is commonly the situation for ectopically expressed genes, wild type males are mutagenized and crossed to females carrying the mutant allele to be modified.

The progeny of the mutagenized and crossed flies that exhibit either enhancement or suppression of the original phenotype are immediately crossed to adults containing balancer chromosomes and used as founders of a stable genetic line. In addition, progeny of the founder adult are retested under the original screening

conditions to ensure stability and reproducibility of the phenotype. Additional secondary screens may be employed, as appropriate, to confirm the suitability of each new modifier mutant line for further analysis.

Although the above-described *Drosophila* genetic modifier screens are quite powerful and sensitive, some genes that participate in the SREBP pathway may be missed in this approach, particularly if there is functional redundancy of those genes. This is because the vast majority of the mutations generated in the standard mutagenesis methods will be loss-of-function mutations, whereas gain-of-function mutations that could reveal genes with functional redundancy will be relatively rare. Methods for systematic gain-of-function screens in *Drosophila* have been developed (Rorth *et al.*, Development (1998) 125:1049-1057). This method is based on a modular mis-expression system utilizing components of the GAL4/UAS system (described above) where a modified P element, termed an "enhanced P" (EP) element, is genetically engineered to contain a GAL4-responsive UAS element and promoter. The resulting transposon is used to randomly tag genes by insertional mutagenesis (similar to the method of P element mutagenesis described above). Thousands of transgenic *Drosophila* strains, termed EP lines, can be generated, each containing a specific UAS-tagged gene. This approach takes advantage of the preference of P elements to insert at the 5'-ends of genes. Consequently, many of the genes that are tagged by insertion of EP elements become operably fused to a GAL4-regulated promoter, and increased expression or mis-expression of the randomly tagged gene can be induced by crossing in a GAL4 driver gene.

Systematic gain-of-function genetic screens for modifiers of phenotypes induced by mutation or mis-expression of an SREBP pathway gene can be performed by crossing several thousand *Drosophila* EP lines into a genetic background containing a mutant or mis-expressed SREBP pathway gene, and further containing an appropriate GAL4 driver transgene. The progeny of this cross are then analyzed for enhancement or suppression of the original mutant phenotype as described above. Those identified as having mutations that interact with the SREBP pathway can be crossed further to verify the reproducibility and specificity of this genetic interaction. EP insertions that demonstrate a specific genetic interaction with a mutant or mis-expressed SREBP pathway gene, have a physically tagged a new gene which can be identified and sequenced using PCR or hybridization screening methods, allowing the isolation of the genomic DNA adjacent to the position of the EP element insertion.

BODIPY-fatty acid conjugates for determining lipid content of nematodes

Because defects in the SREBP pathway can result in abnormal metabolism of lipids, a method for readily identifying mutant model organisms that exhibit abnormalities in lipid metabolism would be beneficial. Existing methods for assessing lipid content in nematodes include the use of non-vital stains such as Sudan Black (Kimura *et al.*, Science (1997) 277:942-6).

However, the drawbacks of these techniques are that the nematodes must be fixed prior to staining. Fixation can introduce artifacts, making an accurate assessment difficult, and furthermore, kills the animals making it impossible to carry out further genetic analysis on the fixed animals. In order to avoid these problems associated with fixing nematodes, certain vital stains were tried that are routinely used for staining lipid in cultured cells such as Nile Red (Greenspan *et al.*, J Cell Biol, (1985) 100:965- 973). However, it was found that these dyes tended to result in background fluorescence of gut granules which are auto-fluorescent organelles of the intestinal epithelial cells that are thought be to lysosomes. In many cases, these fluorescent vital stains appeared to be concentrated in gut granules, enhancing their fluorescence and causing difficulty in accurately measuring the fluorescence due to lipid droplet staining in the intestine. Accordingly, the invention provides an improved method for measuring lipid storage in live nematodes. It has been found that BODIPY® dyes conjugated to fatty acids (e.g. BODIPY® FL C12 (4,4-difluoro-5,7-dimethyl- 4-bora-3a,4a-diaza-s-indacene- 3-dodecanoic acid), and C1-BODIPY® 500/510 C12 (4,4-difluoro-5-methyl-4-bora- 3a,4a-diaza-s-indacene- 3-dodecanoic acid) Molecular Probes, Eugene, OR) concentrate in lipid droplets in the intestines of living nematodes. These dyes do not have the drawbacks associated with other vital dyes because, in addition to clearly staining and fluorescing in lipid droplets in the intestine, they quench the background fluorescence due to the gut granules. Thus, the invention provides a method of using BODIPY®-fatty acid conjugates to stain live nematodes for determining the relative and absolute lipid content in response to changes in metabolic conditions brought on by a) changes in genetic backgrounds including mutations in genes essential for control of metabolic processes, b) changes in environmental conditions such as food sources, temperature, and crowding conditions, and c) different developmental states including the dauer larva. This method is particularly valuable in uses that involve genetic screens and compound screens based on changes in metabolic processes such as the SREBP processing pathway, among others. The method allows considerable increases in accuracy of lipid quantification in vivo over the use of other fluorescent lipophilic stains, making automated sorting of the nematodes based on fluorescence feasible.

BODIPY® conjugates have previously been used to study (1) lipid content in the surface membrane of *Shistosoma mansoni* worms (Redman and Kusel, Parasitology (1996) 113(2):137- 143), (2) lipid endocytosis in cultured mammalian fibroblasts (Pagano and Chen, Ann N Y Acad Sci (1998) 845:152-160), (3) lipid trafficking between the Golgi apparatus and plasma membrane of cultured mammalian fibroblasts (Pagano *et al.*, J. Cell. Biol (1991) 113(6):1267-1279), (4) fatty acid transport by *Saccharomyces* (Faergeman et al., J. Biol. Chem (1997) 272(13):8531-8538) and (5) distribution of ivermectin in muscle vesicle membranes of *Ascaris suum* (Marin and Kusel, Parasitology (1992) 104(3):549-555). However, these prior uses of BODIPY® conjugates do not suggest the applicability of BODIPY® conjugates, and in particular, BODIPY® fatty acid conjugates, for quantification of lipid storage in nematodes. Moreover, the fact that BODIPY® fatty acid conjugates quenches background fluorescence from lysosomes, providing for more accurate quantification, is an unexpected and important advantage provided by the invention that permits large-scale, automated sorting of animals based on fluorescence.

BODIPY®-fatty acid conjugates can be used to stain nematodes of different genetic backgrounds for use in genetic screens, both *de novo* screens for mutations affecting lipid content of whole nematodes and modifier screens for mutations that change lipid accumulation in mutant nematodes (for example, the insulin receptor (*daf*-2) or the SREBP homolog (*ceSREBP*) nematodes). The intestines of the nematodes can be visually examined for lipid content under a fluorescent microscope and mutant animals can be subsequently propagated for cloning purposes. This method can be used in conjunction with automatic flow sorter technology to rapidly separate large numbers of living nematodes by lipid content. This would be useful either for automated high throughput genetic screening or for large scale automated separation of dauer larvae from other developmental stages. Additionally, the method can be used to determine changes in lipid accumulation in nematodes exposed to inhibitory compounds that might serve as therapeutic agents for the control of diabetes, obesity, lipid storage diseases, or other human or animal diseases. A test compound can be administered to a nematode by direct contact, ingestion, injection, or any suitable method and changes in lipid content of the nematode or its progeny are observed. Further, the method is applicable to reverse genetic screening using inhibitory RNA. For example, nematodes could be exposed to combinations of large numbers of RNAs in 384-well plates and screened for changes in lipid content mediated by RNAi using fluorometry or direct visual observation.

EXAMPLES

The following examples show how the nucleic acid sequences of SEQ ID NOs 1, 3, and 5, and 7 were isolated, and how these sequences, and derivatives and fragments thereof, as well as other SREBP pathway nucleic acids and gene products can be used for genetic studies to elucidate mechanisms of the SREBP pathway as well as the discovery of potential pharmaceutical or pesticidal agents that interact with the pathway. As used herein, all *C. elegans*-derived gene sequences are designated by the letters "ce" in front of the gene sequence. Likewise, all *Drosophila*-derived gene sequences are designated by the letter "d" in front of the gene sequence.

EXAMPLE 1: CLONING OF C. ELEGANS SREBP

The *C. elegans* genomic database was searched with the protein sequence of the human SREBP-1, SREBP-2, and *Drosophila* SREBP homologue, HLH106, using the TBLASTN search tool (Altschul et al., *supra*). One *C. elegans* open reading frame showed significant homology with all three of the above SREBP proteins. This homology extends throughout much of the SREBP protein sequences, excluding the C-terminal part of the gene. The *C. elegans* open reading frame was located on two overlapping clones on the right arm of chromosome III (Y47D3 and H10N23). At the time of the search, there were no previous annotations, gene predictions, nor candidate mutants that mapped to this region that would suggest previous identification of this open reading frame as an SREBP-related gene.

Using BLAST analyses (Altschul *et al.*, *supra*) and the GENSCAN Genefinder program (Burge and Karlin, J. Mol. Biol. (1997) 268(1):78-94) , a predicted exon/intron structure for the C. elegans SREBP-related gene (ceSREBP) was generated. This *C. elegans* homologue of SREBP cDNA was cloned in order to validate its existence as an expressed mRNA, and to determine the cDNA and protein sequence for the elucidation of ceSREBP function. Moreover, cloning of ceSREBP was a prerequisite for future genetic manipulations that require knowledge of the sequence, such as RNAi experiments, generation of misexpression constructs, isolation of Tc1 insertion or chemical deletion mutants, etc.

The ceSREBP was cloned by PCR from a mixed-stage, 1st strand cDNA pool that was synthesized from poly-A+ RNA using the NotI primer/adapter (Life Technologies, Gaithersburg, MI). The N-terminal and C-terminal ends of ceSREBP were cloned using gene-specific internal primers and non-specific primers at 3' and 5' ends. The 5' and 3' non-specific primers were, respectively the Splice Leader 1 ("SL1") sequence (Shakes and Epstein, *supra*) and a modified NotI primer/adaptor (Life Technologies). Internal primers were made to regions of high homology

according to the GENSCAN prediction for ceSREBP, that were also predicted by the ACEDB Genefinder (Richard Durbin and Jean Thierry Mieg (1991-present), A *C. elegans* Database; documentation, code and data available from anonymous FTP servers at lirmm.lirmm.fr, cele.mrc-lmb.cam.ac.uk and ncbi.nlm.nih.gov). They were designed to amplify the ends of the cDNA, not the full-length cDNA. Once the end sequence was known, the full-length cDNA was amplified in overlapping N-terminal and C-terminal parts using gene-specific primer pairs. Amplifications were performed using Expand™ High Fidelity buffers and enzyme mixes essentially according to manufacturer's recommendations (Roche, Summerville, NJ). PCR amplicons were subcloned into the PCR2.1 intermediate vector (Invitrogen, San Diego, CA) and sequenced using the Big Dye™ dye-terminator sequencing kit and the ABI377 sequencer (ABI, Foster City, CA). Sequence analysis was done with the Sequencher program (Gene Codes Corporation, Ann Arbor, MI). Sequencing primers were designed using the Oligo 4.0 program (Molecular Biology Insights, Inc.; Cascade, CO) and were selected from the predicted gene sequence and, subsequently, from sequence of newly obtained fragments.

Sequence information obtained from the initial amplified fragments allowed subsequent amplification overlapping fragments that represented the entire cDNA. Any sequence discrepancies, either with the available genomic sequence or among different subclones of the same sequence, were resolved by sequencing multiple, independently isolated subclones of the given fragment.

A single contiguous sequence ("contig") was assembled that shared sequence identity with the YAC sequence Y47D3 (GI3646936) from which gene predictions were made. It was mostly identical to the gene prediction in this region but contained one exon not included in the gene prediction. The contig contained a single open reading frame (ORF), initiation and termination codons, and a likely poly-adenylation signal. BLAST analysis of this contig against GenBank sequences, showed highest homology to other SREBPs.

The cDNA sequence of the ceSREBP gene, SEQ ID NO:1 is shown in Figure 2. The cDNA is 3419 nucleotides long. This full-length clone contained a single open reading frame with an apparent translational initiation site at nucleotide position 24 and a stop signal at nucleotide position 3365. The predicted polypeptide precursor is 1113 amino acids. Additional features include an acidic domain at about nucleotides 24 to 233 (amino acid residues 1 to 69); a possible second acidic domain at about nucleotides 987 to 1040 (amino acid residues 321 to 338); a basic Helix-loop-helix domain at about nucleotides 1089 to 1286 (amino acid residues 355 to 421); a first transmembrane domain at about nucleotides 1455 to 1514 (amino acid residues

477 to 497); and a second transmembrane domain at about nucleotides 1653 to 1706 (amino acid residues 543 to 561).

A BLAST analysis against the Y47D3 clone which has a total of 351,956 nucleotides, revealed 12 regions of Y47D3 which share sequence identity with SEQ ID NO:1, as shown in Table I.

TABLE I

Base # of SEQ ID NO:1	Base # of Y47D3	% Sequence Identity	Exon #
1-80	179,410-179,331	100	1
81-213	178,918-178,786	100	2
214-523	178,528-178,218	100	3
527-632	177,448-177,338	96	4
633-1052	177,286-176,864	100	5
1053-1288	176,520-176,285	100	6
1289-1482	175,768-175,568	100	7
1483-2011	175,523-174,994	100	8
2012-2408	174,687-174,288	100	9
2409-2636	174,228-174,001	100	10
2637-2790	173,954-173,801	100	11
2791-3151	155,054-154,694	100	12
3152-3397	154,638-154,393	100	13

An alignment of the predicted protein sequence (SEQ ID NO:2) against the human and Drosophila SREBP proteins was performed. Amino acid residues 353 to 423 of SEQ ID NO:2 share 45% and sequence identity and 77% sequence similarity with amino acid residues 281-351 of *Drosophila* SREBP (Theopold *et al.*, *supra*; GI079656). Amino acid residues 466 to 826 of SEQ ID NO:2 share 28% sequence identity and 47% sequence similarity with human SREBP2 (GI1082805).

The presence of other gene and protein sequences bearing significant homology to the full-length ceSREBP was further investigated using the BLAST family of computer programs against public databases. The following amino acid sequences were the most similar: SREBP-1, Chinese Hamster (GI 1083186); SREBP-1, Cricetulus griseus (GI 516003); Sequence 54 from patent US 5527690 (GI 1610915); SREBP2 precursor, human (GI 1082805); SREBP-2, Homo sapiens (GI 451330); SREBP2 precursor, Chinese hamster (GI 1083185); Sequence 38 from patent US 5527690 (GI 1610908); SREBP-1, Homo sapiens (GI 409405); SREBP-2,

Cricetulus griseus (GI 551506); Transcription factor ADD1, Rat (GI 540006); and HLH106, *Drosophila Melanogaster* (GI 107965).

Subsequent to the above analysis, a Genefinder prediction of the ceSREBP protein was entered into the Genbank database, which is 100% identical to SEQ ID NO:2, and is designated GI 3881008.

EXAMPLE 2: ceSREBP EXPRESSION ANALYSIS

Expression of ceSREBP was assayed using a transcriptional reporter system in which the putative promoter/enhancer region of ceSREBP was fused to GFP. To determine how much genomic sequence to include in the reporter construct, the Y47D3 contig containing the N-terminal region of the ceSREBP cDNA and ~25 kb upstream of the ceSREBP initiation codon, was analyzed using ACEDB Genefinder and GENSCAN programs (Burge and Karlin, *supra*). There were no known genes within this region, and no predicted genes reported by either program. Of the two predicted genes within ~8 kilobases (kb) of ceSREBP, one, ~5 kb upstream of ceSREBP showed limited homologies by BLAST analysis to *C. elegans* expressed sequence tags (ESTs). A genomic fragment of ~4.5 kb was chosen as the putative promoter/enhancer region.

The promoter enhancer fragment was amplified by PCR. PCR primers were designed to amplify the ~4.5 kb genomic fragment, including the first few amino acids of ceSREBP. Restriction sites were included in the primers to facilitate sub-cloning into the GFP reporter vector pPD117.01 (from the laboratory of Dr. Andrew Z. Fire (Fire Lab)), Carnegie Institution of Washington, Baltimore, MD) in an in-frame translational fusion to GFP. The sense primer, nucleotides 71,242-71,265 of Y47D3 (GI:3646936), contained an AscI site; the antisense primer, nucleotides 66,719-66,747 of Y47D3, contained a KpnI/Asp718 site:

PCR was performed in duplicate to provide two independently amplified promoter fragments for independent confirmation of the expression pattern and thus control for sequence errors introduced by PCR. The Klentaq enzyme and buffer systems were used essentially according to manufacturer's protocols. From each original PCR reaction, a single clone that contained the expected insert was identified. The putative promoter/enhancer fragments were ligated into AscI and Asp718 restriction sites in pPD117.01 to create "ceSREBP::GFP."

ceSREBP::GFP was injected into N2 worms using standard protocols for *C. elegans* transformation (Epstein and Shakes, *supra*) at a concentration of 10 µg/ml plus 100 µg/ml pRF4 rol-6(d) transformation marker, and stable lines displaying the roller phenotype were established.

Expression analysis was done with a fluorescence microscope (Axioplan™, Zeiss, Thornwood, NY). By GFP expression analysis, ceSREBP is first expressed weakly in embryonic gut cells at the time of gut cell polarization, which marks the beginning of differentiation. There is strong fluorescence by the "bean stage" which persists in all intestinal cells throughout embryogenesis and at all larval and adult stages. There is also weak fluorescence in the pharynx. Because there is high specificity of expression of ceSREBP in intestinal cells, the ceSREBP promoter, contained within nucleotides 66,719-71,265 of Y47D3 (GI:3646936), has utility as a tissue specific promoter that can be operably linked to heterologous sequences, such as marker genes and/or genes of interest. Thus, the ceSREBP promoter can be used for studying biochemical pathways within the intestine of *C. elegans*.

EXAMPLE 3: RNA INTERFERENCE (RNAI) OF *C. ELEGANS* SREBP, S2P and SCAP

RNAi experiments were performed with *C. elegans* SREBP pathway genes in order to generate loss-of-function phenotypes that could help elucidate the function of the SREBP pathway in *C. elegans*.

Methods:

PCR was carried out on *C. elegans* sequences for SREBP (SEQ ID NO:1) and S2P (Rawson et al., supra; GI1559384), and a Genbank sequence (GI3875380), that is annotated as having HMG-CoA reductase homology, and additionally has been determined to have homology to the human SCAP protein. Accordingly, GI3875380 is referred to herein as ceSCAP. Fragments of between 0.3kb to 1.2kb were produced in regions of interest. In general, mixtures of fragments that individually corresponded to different regions of the genes and together covered most of the central region of each gene were injected. For ceSREBP, these fragments spanned a mid-region of ~1.4 kb, excluding the amino-terminal ~ 0.6 kb and the carboxy-terminal ~1.4 kb. Similarly, ceS2P dsRNA fragments excluded the amino terminal ~0.7 kb and carboxy-terminal ~0.3 kb and covered the central region. ceSCAP fragments covered most of the coding sequence, excluding the amino-terminal ~0.4 kb and the carboxy-terminal ~1kb. PCR was carried out with 0.5μM each primer and 0.4 μg genomic DNA using the Expand™ PCR Kit (Roche) at 94°C 1 min 15 sec; and 35 cycles of: 94°C 15 sec, 57°C 45 sec, and 72°C 1 min.

A small fraction of each reaction was run on a gel to assure that the PCR worked. The rest of each reaction was precipitated and then re-suspended in RNase-free water, to serve as the template for production of sense and antisense RNAs. Sense and antisense RNA were transcribed together from the DNA template using T7

RNA polymerase (Promega, Madison, WI; RNA production kit, Cat#1300) following the manufacturer's protocol. The resulting RNA samples were ethanol-precipitated and resuspended in 20 µl of RNAse- free TE (10mM tris, 1mM EDTA), followed by 10 µl of RNase free 3X IM annealing buffer (20mM KPO4 pH7.5, 3mM KCitrate pH 7.5, 2% PEG 6000). The reactions were incubated at 68°C for 10 minutes and then at 37°C for 30 minutes to anneal the sense and antisense strands.

Injection volumes were selected to deliver $0.5x10^6$ to $1x10^6$ molecules of RNA. Injections were delivered to the gonads or the intestinal cavity of *C. elegans*, and were carried out using the methods of Fire *et al.* (Development (1991) 113:503-514).

For germline RNAi, adult animals were microinjected with RNA into either the gonad or intestine using a glass needle mounted on a Medical Systems Corp. (Holliston, MA) PLI-90 injector. For RNAi of larvae, wild type L1 larvae were isolated by first collecting embryos from gravid adults by digestion in 1.25% sodium hypochlorite, 0.25M potassium hydroxide, and then allowing the embryos to hatch overnight in M9 buffer. Equal volumes of larvae in M9 buffer and RNA were mixed in wells of microtiter plates, incubated for 24 hours at 15°C, and then transferred to standard nematode growth plates.

For visualization of lipid, some of the worms were washed off a plate using M9 buffer (per liter: 30gr Na_2HPO^4, 15g KH_2PO^4, 2.5g NaCl, 5g NH_4Cl), collected by centrifugation, and resuspended in a 2ng/ml solution of BODIPY™.FL.Cl2; stock solution is 1mg/ml in ethanol) prepared in M9 buffer. The worms were placed on a benchtop shaker overnight at room temperature to absorb the dye. Images were captured using a fluorescence microscope (Axioplan™, Zeiss, Thornwood, NY) the next day.

Results:

ceSREBP RNAi

Germline ceSREBP RNAi produced several visible phenotypes in the progeny of the microinjected animals. The gross phenotype was a fully penetrant larval arrest. Arrested larvae appeared to be at the L2 stage based on gonad and cuticle morphology, although their length was more similar to that of L1 stage larvae. Arrested larvae remained motile and feeding for several days at 20°C before dying. Their intestine appeared paler, or less darkly pigmented, than wild type, and this is referred to as the "pale intestine" or "Pin" phenotype.

Morphological defects in ceSREBP RNAi larvae (L1 and L2 stages) were confined to the intestine, where ceSREBP appeared to be primarily expressed, and

specifically affected three cytoplasmic structures in intestinal cells. First, there was a dramatic reduction in the number and average size of pigmented droplets in the intestine. This reduction of pigmented droplets seemed to account for the Pin phenotype observed at low magnification. These droplets likely contain lipid since they stained with dye-labeled fatty acid (BODIPY™-dodecanoic acid) and their number in various developmental stages and mutants correlated with the level of staining with the dye Sudan black in fixed animals.

These observations indicated that ceSREBP is required for formation and/or maintenance of lipid droplets in the intestine, the main lipid storage organ of *C. elegans*. Second, the gut granules appeared larger and more birefringent than in wild type. Third, many variably sized vesicles appeared in the intestine. These vesicles were spherical and transparent; similar vesicles are only rarely observed in wild type larvae. The vesicles in ceSREBP RNAi larvae were usually each associated with a gut granule, and they showed autofluorescence similar in color and intensity to that of gut granules. Since gut granules are thought to be lysosomal structures, the abnormal vesicles in ceSREBP RNAi larvae may also have been lysosomal in origin. The larval arrest and morphological defects in the intestine described above were also observed in mutant larvae homozygous for the ceSREBP partial deletion allele "ep79 (see Example 4), " suggesting that germline RNAi reproduced the zygotic null phenotype.

ceSREBP RNAi of larvae at the L1 stage resulted in apparently normal development through the L2 stage, with approximately normal accumulation of intestinal pigmented droplets. However, most larvae arrested at the L3 or L4 stage and failed to maintain their droplets. Arrested larvae, as well as many fully developed adults, showed the Pin phenotype and had a thinner body than normal. The number and size of pigmented intestinal droplets was greatly reduced, as observed in earlier stages for germline RNAi. The finding that the Pin phenoptype could be induced by RNAi treatment after terminal differentiation of the intestine indicates that the phenotype is unlikely to be caused by a developmental defect in the intestinal cells. Rather, ceSREBP may be required continuously for proper functioning of the intestine. The pale, thin appearance of ceSREBP RNAi larvae and adults was similar to that of starved animals; however, the RNAi animals display foraging behavior and pump in bacteria through the pharynx into the intestine. These observations suggested that ceSREBP RNAi larvae were defective in digesting and/or metabolizing food. ceSREBP RNAi larvae showed greater dispersal away from the food source than wild type, possibly because they could not derive nutrients from the bacteria. Gut granules of the arrested larvae and adults were often larger and more birefringent that normal.

Adults that displayed the Pin phenotype had fewer embryos than normal in their uterus, suggesting reduced fecundity, and some of the embryos showed variable developmental defects. Finally, Pin adults often contained large, transparent vacuoles in the anterior half of the intestine. These vacuoles were distinct from the abnormal vesicles observed in larvae, since the vacuoles were irregularly shaped and not autofluorescent, although their origin remains unidentified. ceSREBP RNAi of larvae at the L2 stage resulted in the same defects as L1 treatment, but mainly in later stages of development. Most animals arrested at the L4 stage or displayed the adult defects.

We also performed a RNAi feeding experiment. A DNA vector in which a portion of the ceSREBP cDNA was inserted between two T7 promoters was constructed and introduced into *E. coli* that express the T7 polymerase. The fragment contained most of the coding sequence except the amino-terminal ~600 nucleotides. The bacteria generated both sense and antisense transcripts, and thus produced dsRNA of ceSREBP. When fed to wild type nematodes, these E. coli produced the same phenotypes obtained from RNAi treatment of L2 larvae, namely, pale intestine, reduced body size, and reduced fecundity.

ceS2P RNAi

Germline RNAi of the ceS2P resulted in apparently normal development through the adult stage. However, adults showed a fully penetrant phenotype, exhibiting all the defects observed for ceSREBP larval RNAi except larval arrest. Specifically, the adult phenotype included a small, thin body, pale intestine associated with few lipid droplets, abnormally large and birefringent gut granules, large vacuoles in the anterior intestine, fewer embryos in the uterus, and variable developmental defects in some of the embryos. The gut granule defects seemed more pronounced than observed for ceSREBP RNAi. The striking similarity of the RNAi phenotypes for ceS2P and ceSREBP strongly suggested that these two genes function in a common genetic pathway. The lack of effect of ceS2P RNAi on larval development may indicate functional redundancy with an unidentified gene or reduced potency of RNAi for ceS2P compared to ceSREBP.

ceSCAP RNAi

Germline RNAi of the SCAP homologue generated a phenotype similar to ceS2P RNAi in less than 10% of adults. Defective adults displayed a pale intestine, small and thin body, few embryos in the uterus, and slightly more birefringent gut granules. Germline RNAi of both ceS2P and ceSCAP together produced a fully penetrant phenotype indistinguishable from ceSREBP germline RNAi. This

phenotype included L2-L3 larval arrest, pale intestine associated with few or no intestinal lipid droplets, and abnormally large and birefringent gut granules. These results suggested that both the ceS2P and ceSCAP homologues function in the ceSREBP genetic pathway at all larval and adult stages. If RNAi of ceS2P or ceSCAP produced the null phenotype for these genes, then there must exist other gene activities that can partially substitute for their functions, presumably in proteolytic cleavage at site 2 and 1 analogues, respectively, of ceSREBP.

EXAMPLE 4: TC1 TRANSPOSON MUTAGENESIS

The goal of this set of experiments was to produce loss-of-function mutations in genes of interest in order to understand the function of their wild-type counterparts.

Library preparation

A Tc1 transposon insertion library comprising 3 sets of 960 cultures was constructed according to published protocols (Zwaal *et al.*, *supra*, and Plasterk, *supra*).

Library screening

The library was screened in individual tiers, each library having three tiers, with each tier composed of 1,000 lysates or ~200,000 haploid genomes. Lysates were pooled according to the published protocol. A first dimension screen involved PCR on 8 samples of pooled DNA from ten 96-well plates. A second dimension screen was used to determine which of the ten 96-well plates contained the desired mutant (involved screening of 10 DNA pools). A third dimension screen was used to determine the "address" of a particular mutant (*i.e.*, in which column and row a particular mutant resided - via screening of 12 individual lysates from a single row). First dimension reactions were done in quadruplicate; second and third were done in triplicate.

Two rounds of PCR were performed, each with one gene-specific primer and and oneTc1-specific primer. The second PCR reaction was designed to be a nested reaction. Accordingly, two nested pairs of Tc1 primers were used: one pair pointing outward from the left of the transposon, and the other pair pointing outward from the right (these primer pairs are described in the references cited above). The gene-specific primers are all contained within SEQ ID NO:1 or intervening intron sequence contained within Y47D3 (see Table I in Example I).

The first and second round PCR for each dimension was performed in 15 µl total volume using the following in each reaction: 1X PCR buffer provided by the manufacturer (Perkin-Elmer), 1.5 mM $MgCl_2$, 0.2 mM dNTPs, 0.5 µM each of the

Tc1 and the gene-specific primer, 0.5 units of Taq Polymerase (Perkin-Elmer); and H_2O to 13 µl for the first round reactions, and to 15 µl for the second round

First and Second dimension: 2 µl of 1:20 diluted DNA was added; 1:10 DNA diluted was added to the third dimension reactions. A small amount of first round reaction was transferred to the second round using a pin replicator. PCR cycling conditions were: 94°C for 3 minutes; then 94 °C for 40 seconds, 58°C for 1 minute, 72°C for 2 minutes for 35 cycles; then 72°C for 2 minutes.

Identification of insertion animals:

Four confirmed Tc1 insertions were found in the ceSREBP gene, in introns 2, 5, 7, and 8, which follow exons 2, 5, 7 and 8, respectively, in the genomic sequence in Table I in Example I. All addresses are from Tier 1 of the Tc1 library described above.

The "6D2" address, located downstream of the two predicted transmembrane domain coding regions was chosen for further analysis based on its relatively central location within the SREBP gene.

Nematodes were recovered from a frozen stock representing the 6D2 address. In order to identify a nematode carrying the insertion, individual surviving nematodes were transferred to individual plates. After progeny from these nematodes were present on the plates, the parent nematodes were transferred into individual wells of a 96-well plate that contained 5µl of nematode lysis buffer (100 mM KCl, 20 mM Tris-HCl pH 8.3, 5 mM MgCl2, 0.9% Nonidet P-40, 0.9% Tween-20, 0.02% gelatin, and 400 µg/ml proteinase K). The nematodes were lysed in a PCR machine at 60°C for one hour, followed by 95°C for 15 minutes. 18 µl of a PCR master mix then was added to the crude lysates (to give ~20 µl total reaction volume, assuming evaporation of a portion of the lysate); this mix contained 1X reaction buffer (Perkin-Elmer), 1.5 mM $MgCl_2$, 0.2 mM each dNTP, 0.5 µM each gene-specific primer, 0.5 units Taq polymerase, and dH_2O to 18 µl per reaction

The PCR reactions were cycled using a program identical to that used for screening the library for the insertions described above. Subsequently, a second round of PCR was performed using the same conditions and primers described above for the insertion screen, after transferring a small amount of the first round reaction to the second round master mix using a pin replicator. Reactions were analyzed on agarose gels for insertion products identical in size to those observed in the original screen for insertions.

Using this PCR-based screen, a population of nematodes was obtained that was homozygous for the insertion. However, since this Tc1 insertion was in an intron, and since Tc1 elements are often completely removed along with the intron during splicing of the pre-mRNA, the insertion was unlikely to affect gene function.

Consequently, this insertion population was used to identify a deletion in the ceSREBP gene by imprecise excision of the Tc1 element (as described above).

Identification of a Tc1-mediated deletion

In order to obtain a Tc1-mediated deletion in the ceSREBP gene, a small library consisting of 244 cultures of 6D2 insertion nematodes was generated. To create the library, ~5-10 nematodes homozygous for the 6D2 insertion were seeded onto individual plates. After these nematodes had grown, reproduced, and consumed all of the bacteria on these plates, triplicate lysates representing these cultures were created. A sample of nematodes from each plate was collected by washing the plate with distilled water, and placing the nematodes washed from each plate in one well of a 96-well plate. This was repeated two additional times to create a triplicate set of lysates. Nematodes were lysed by addition of an equal volume of lysis buffer followed by incubation at -80°C for 15 minutes, 60°C for 3 hours, and 95°C for 15-30 minutes.

Deletion screening was carried out using a PCR-based approach similar to that used for insertion screening, both of which have been described previously (Zwaal et al., *supra*; and Plasterk, *supra*). Two sets of gene-specific primer pairs were chosen for carrying out a nested PCR strategy such that an outside set was used for the first round of PCR and an inside set was used for the second round of PCR. The second round of PCR was performed to achieve greater specificity in the reaction. The primer sets chosen were ~3.2 kb apart in the ceSREBP genomic sequence (within the typical range for Tc1 deletion screening) flank either side of the Tc1 insertion in the 6D2 population.

The first round PCR reactions were performed using 2 µl of lysate from two of the three sets of lysates, with reactions carried out in a 96-well plate. PCR was done as described above for identification of insertions.

The reactions were carried out in duplicate using the following cycling parameters: 94°C for 3 minutes, then 35 cycles of the following: 94°C for 40 seconds, 55°C for 1 minute, and 72°C for 1 minute. The second round of PCR was essentially the same. A small amount of first-round reaction products was transferred to the second-round reaction mixtures using a 96-pin replicator.

Products of the second round of PCR were analyzed by electrophoresis in 1% agarose gels. A potential deletion product was observed in both of the reactions, and the putative positive lysate was re-tested by performing duplicate reactions using the relevant lysate from all 3 sets of the library (for a total of six reactions) in two rounds of PCR as described above. The product was gel purified and sequenced directly to

confirm the presence of the desired deletion. In addition, in order to confirm that the deletion product obtained was specific for the SREBP region (*i.e.* not an artifact of the PCR), an additional primer set was used in two rounds of PCR. A separate set of reactions was performed using all three lysates and one of the two original primer pairs. This primer set was chosen such that the PCR product generated would be ~100-300 base-pairs different in size from the original deletion product, resulting in a noticeable shift in size from the original product when analyzed on 1% agarose gels. This part of the screening procedure is termed the "specificity test". Using this procedure to screen the 244 lysates from the 6D2 insertion library with the primers listed above, one deletion of ~2.2 kb within the ceSREBP genomic region was identified, and confirmed by the specificity test and by sequence analysis. This deletion begins within intron 6, and ends within exon 9 of the ceSREBP gene. This partial deletion allele was named ceSREBP (ep79).

Identification of deletion animals

Following the identification the deletion, 192 individual nematodes from the relevant plate were transferred onto separate plates. When progeny were present on the plate, the parent nematodes were placed into buffer in 96-well plates and lysed as described above. PCR was performed using the primers that identified the deletion. One animal that carried the deletion was identified.

Analysis of mutant phenotypes

Prior to analysis of the SREBP deletion animals, animals carrying the SREBP deletion were out-crossed ten times to a wild-type (N2/Bristol) strain in order to remove unrelated mutations induced by Tc1 elements in the original mutator strain from which the insertion and deletion animals were isolated. Throughout the out-crossing procedure, the ceSREBP deletion was followed and maintained by analyzing progeny of these crosses by PCR, using the same primers and conditions used for the deletion screen above.

The deletion mutation was placed in *trans* to a balancer chromosome and maintained as a heterozygous strain. Reduction or elimination of function mutations often recapitulate phenotypes observed by RNAi, and ceSREBP RNAi resulted in larval arrest. Accordingly, homozygous deletion mutants would not be able to propagate themselves if the mutation produced a larval arrest phenotype.

The out-crossed and balanced strain was analyzed for any mutant phenotypes due to the SREBP deletion. Approximately 25% of the progeny derived from heterozygous SREBP deletion animals (which would correlate to presumptive deletion homozygotes) displayed the same phenotypes produced by ceSREBP RNAi, described

in Example 3 above. These phenotypes included early larval arrest, reduced pigmentation as a result of fewer lipid droplets in the intestine, and accumulation of fluid-filled vesicles.

EXAMPLE 5: CLONING OF DROSOPHILA S2P

Using BLAST, two EST clones from the Berkeley Drosophila Genome Project (BDGP), LD11632 (GI2044683) and LD14421(GI2151648) were found to have homology with hamster S2P (GI2745731). The sequences were contained in two P1 clones D379 and D380 (GI3893020). Primer walking based on these sequences was used to obtain the full-length DNA sequence. Several more sequencing reactions were performed to produce a complete and unambiguous coverage of the gene, which is referred to herein as Drosophila S2P (dS2P). A full length clone (SEQ ID NO:3) was identified that contained a single open reading frame with an apparent translation start site at nucleotide position 219, and a stop signal at nucleotide position 1745. The predicted polypeptide precursor is 508 amino acids long (SEQ ID NO:4). A search of the PFam and PROSITE databases (Sonnhammer et al., Genomics (1997) 46:200-216; Bairoch et al. NAR (1991) 19 Suppl:2241-2245; and Hofmann et al., NAR (1999) 27:215-219) revealed seven transmembrane domains and a PDZ domain. The transmembrane domains are located at approximately amino acid residues 4 to 20 (TM1), 82-98 (TM2), 143-159 (TM3), 163-179 (TM4), 208-224 (TM5), 428-444 (TM6) and 478-494 (TM7). The putative PDZ domain is located at approximately amino acid residues 215-285.

The presence of other gene and protein sequences bearing significant homology to Drosophila S2P (Fig.2, SEQ ID NO.4) was investigated using the BLAST family of computer programs (Altschul et al., *supra*). The following amino acid sequences were the most similar: S2P *Homo sapiens* (GI2745733); S2P *Cricetulus griseus* (GI2745731); SP2 metalloprotease, *Homo sapiens* (GI4164134 and GI4164135); putative protein *Arabidopsis thaliana* (GI2982448); conserved protein *Methanobacterium thermoautotrophicum* (GI2622476); and Orf c04034 *Sulfolobus solfataricus* (GI1707806). The most homologous sequence was human S2P (GI2745733) which shared 9 contiguous amino acids at positions 201-207 of SEQ ID NO:4. Amino acids 127 to 501 of SEQ ID NO:4 shares 32 % sequence identity with amino acids 148 to 515 of GI2745733.

EXAMPLE 6: CLONING OF DROSOPHILA SCAP

The Drosophila SCAP homologue (dSCAP) identified herein, was cloned by PCR based on sequence from a gene prediction and from 5' RACE. BLAST analysis

of the hamster SCAP (GI1675220) revealed a genomic P1 clone, DS06954, with regions of high homology. GENSCAN gene finder analysis of this P1 predicted a cDNA that included these homologous regions and was partially covered by ESTs. dSCAP was cloned in overlapping N-terminal and C-terminal fragments.

N-terminal sequence not represented within the gene prediction was obtained by RACE from embryo cDNA prepared with Marathon system (Clontech). A short N-terminal fragment was amplified using non-specific primer AP1 to the Marathon adaptor and an antisense primer contained within the 5' EST. Amplification was performed with Klentaq™ enzyme and buffers, essentially according to manufacturer's recommendations.

The major ~0.7 kb PCR product was cloned into the pCRII shuttle vector (Invitrogen) and completely sequenced using M13 forward and reverse primers, and the start codon was identified. Based on the N-terminal sequence identified, a longer N-terminal and an overlapping C-terminal fragment were amplified.

A full-length contig was identified that contained a single open reading frame with an apparent translational initiation site at nucleotide position 73 and a stop signal at nucleotide position 3786 (SEQ ID NO:5). The predicted polypeptide precursor is 1237 amino acids long (SEQ ID NO:6). Additional features include:

1) Four G-beta (GB) repeat WD domains: GB1 at nucleotides 2509 to 2617, corresponding to amino acid residues 812 to 848; GB2 at nucleotides 3080 to 3196, corresponding to amino acid residues 1005 to 1041; GB3 at nucleotides 3208 to 3325, corresponding to amino acid residues 1045 to 1084; and GB4 at nucleotides 3337 to 3445, corresponding to amino acid residues 1088 to 1124;

2) Six predicted transmembrane (TM) domains. TM1 at nucleotides 991 to 1039, corresponding to amino acid residues 306 to 322; TM2 at nucleotides 1117 to 1165, corresponding to amino acid residues 348 to 364; TM3 at nucleotides 1180 to 1228, corresponding to amino acid residues 369-385; TM4 at nucleotides 1366 to 1414, corresponding to amino acid residues 431-447; TM5 at nucleotides 1753 to 1801, corresponding to amino acid residues 560 to 576; and TM6 at nucleotides 2353 to 2401, corresponding to amino acid residues 760 to 776.

The presence of other gene and protein sequences bearing significant homology to dSCAP (SEQ ID NO:5) was investigated using BLAST (Altschul et al., *supra*) against nucleotide databases. This revealed that dSCAP is covered by two genomic clones from BDGP: DS06954 (P1 D338, GI 4454581), and DS05325 (P1 D340, GI 4454581). The accession number for the two clones is AC007121. Other sequences bearing nucleotide homology with dSCAP are human mRNA for KIAA0199 gene (GI 1228046), and *Cricetulus griseus* SCAP mRNA (GI 1228046).

At the protein level, dSCAP shares homology with the following sequences: *C. elegans* predicted SCAP D2013.8 (GI 642180), *Homo sapiens* KIAA0199 gene (GI 1228047), *Cricetulus griseus* SCAP (GI 1675220), and is similar to the transmembrane domain of HMG-CoA reductase (GI 3875380).

EXAMPLE 7: TRANSGENIC DROSOPHILA MISEXPRESSING SREBP

A constitutively active form of *Drosophila* SREBP (dSREBP.CA) was engineered for misexpression in *Drosophila*, to be used both as a screening or counter-screening reagent and a device to further elucidate the function of SREBP in *Drosophila*. dSREBP.CA contained amino acids 1-448 of dSREBP (SEQ ID NO:8), including the acid blob and bHLH-Zip domain, but terminated just prior to the first membrane-spanning domain, where a stop codon was added. It thus required no proteolytic processing for activation of transcriptional targets.

The construct was cloned into pExPress-UAS. pExPress is a vector designed specifically for misexpression of genes in transgenic *Drosophila* that was derived from pGMR (Hay *et al.*, *supra*). The vector is 9Kb long, and contains the following components: an origin of replication for *E. coli*, an ampicillin resistance gene, P element transposon ends, a *White* marker gene, and an expression unit comprising the TATA region of *hsp70* enhancer and the 3' untranslated region from the *α-tubulin* gene. The expression unit contains a first multiple cloning site (MCS) designed for insertion of an enhancer and a second MCS located 500 bases downstream, designed for the insertion of a gene of interest. dSREBP.CA was cloned into the EcoRI site of the second MCS.

The expression construct was injected into yw *Drosophila* embryos using standard protocols for *Drosophila* transformation (Rubin and Spradling, *supra*). A variety of GAL4 driver lines were used to drive mis-expression of the transgenes. Driver lines *Kruppel*, *Rhomboid*, and 1878 are available from the University of Indiana (http://flybase.bio.indiana.edu) and drive expression in many tissues in 3rd instar larvae, including the gut, fat body and nervous system. Lines T93 and T113 were kindly provided by Tian Xu (Yale University School of Medicine, New Haven, CT) and drive expression primarily in larval fat body.

dSREBP.CA produced a lethal phenotype when expressed by the 1878, *Kruppel*, or *Rhomboid* drivers. Expression via the T93 driver produced a range of phenotypes which included reduced male viability, reduced female fertility, adults with shrunken abdomens and a starved appearance, persistence of the larval fat body in adults, and a short life span. Expression via the T113 driver produced mostly pupal lethality. Most adult survivors were female and displayed the shrunken abdomen

phenotype. These phenotypes, produced by expression primarily in the larval fat body, provided evidence that the dSREBP.CA transgene exerted metabolic effects.

EXAMPLE 8: CLONING OF dS1P NUCLEIC ACID SEQUENCE

The *Drosophila* homologue of mammalian S1P was cloned based on sequence from an EST library that was prepared from tissue from mixed stage embryos, larval imaginal discs, and adult fly heads. mRNA from these tissue was used to construct a random primer library that was normalized using a modification of the method described by Bonaldo *et al.* (Genome Research (1996 6:791-806). The total number of colonies picked for sequencing from the normalized library was 240,000. The reactions were primarily carried out with primer that initiated at the 5' end of the cDNA inserts. Clones that were of biological interest or that could extend assemblies of contiguous sequences ("contigs") were also sequenced from the 3' end. DNA sequencing was carried out using ABI377 automated sequencers and used either ABI FS, dirhodamine or BigDye chemistries (Applied Biosystems, Inc., Foster City, CA).

Analysis of sequences was done using the program "Phred" (Gordon, Genome Res. (1998) 8:195-202). Each sequence was compared to all other fly EST sequences using the BLAST program and a filter to identify regions of near 100% identity. Sequences with potential overlap were then assembled into contigs using the programs "Phrap", "Phred" and "Consed" (Phil Green, University of Washington, Seattle, Washington; http://bozeman.mbt.washington.edu/phrap.docs/phrap.html). The resulting assemblies were then compared to existing public databases and homology to known proteins was then used to direct translation of the consensus sequence. The contig sequences were archived in an Oracle-based relational database (FlyTag™, Exelixis Pharmaceuticals, Inc., South San Francisco, CA).

The predicted full-length dS1P nucleic acid was amplified using PCR and sequenced. Sequences were assembled using Phred/Phrap and analyzed using Consed. This effort resulted in a contiguous nucleotide sequence of 3160 bases in length, encompassing an open reading frame (ORF) of 2979 nucleotides encoding a predicted protein of 993 amino acids. The ORF extends from base 62-3040 of SEQ ID NO:1.

Upon completion of cloning, the sequences were analyzed using the Pfam (Bateman *et al.*, Nucleic Acids Res. (1999) 27:260-262) and Prosite (Hofmann *et al.*, Nucleic Acids Res. (1999) 27(1):215-219) programs to identify motifs in the resulting translations. Four transmembrane domains were predicted at amino acids 1-21, 363-383, 458-478, and 500-520 (corresponding to nucleotides 62-130, 1150-1213, 1435-

1498, and 1561-1624, respectively). Prosite predicted homology to the peptidase_S8 subtilase family (PS00138) at amino acids 154-425 (nucleotides 523-1339).

Nucleotide and amino acid sequences for the dS1P nucleic acid sequence and the encoded protein were searched against all available nucleotide and amino acid sequences in the public databases, using BLAST (Altschul *et al.*, *supra*). At the protein level, the closest homologs predicted by BLAST analysis were the human and hamster S1P sequences (GI 4506775 and GI 7522635) and mouse and rat proteins related to subtilisin (GI4679093 and GI 4679095).

EXAMPLE 9: GENETIC INTERACTIONS OF C. ELEGANS SREBP, INSULIN RECEPTOR, AND TGF-B PATHWAY GENES

Genetic and Phenotypic Analyses:

Loss-of-function and reduction-of function mutations in either the *daf-2*/InR (GI2338417) or *daf-7*/TGF-β (GI1684866) signaling pathways that confer a dauer-constitutive phenotype also produce a dark intestine phenotype associated with increased accumulation of lipid droplets in the intestine (Kimura KD *et al.*, Science (1997) 277:942-946). The dark intestine phenotype can occur in late larvae and adults independent of dauer formation, and appears to be the reciprocal of the pale intestine phenotype cause by reduction of SREBP function and associated with decreased accumulation of intestinal lipid droplets.

We have demonstrated that mutations in *daf-2*/InR, *daf-7*/TGF-β, or downstream components in either pathway can partially suppress reduction of SREBP function. Specifically, *daf-2 (e1370)* (P1465S mutation in kinase domain; Kimura *et al.*, *supra*) adults grown on *E. coli* expressing double-stranded RNA for part of the *C. elegans* SREBP gene produced larvae and adults that showed significantly more intestinal pigmentation and lipid droplets than similarly treated wild-type animals. The accumulation of lipid droplets was intermediate in amount between that of RNAi-treated and untreated wild-type animals. This experiment was performed with the temperature-sensitive mutation *daf-2 (e1370)* at 20°C, a temperature at which most mutant animals develop to adulthood and display a dark intestine phenotype on wild-type *E. coli*. We also constructed a double mutant combination of *daf-2 (e1370)* with the partial deletion mutant ceSREBP *(ep79)*. The *daf-2 (e1370)* mutation partially suppresses the pale intestine phenotype of the ceSREBP *(ep79)* mutant, as well as the small and thin body size phenotype and the small brood size phenotype. The *ep79* mutation deletes the C-terminus of the SREBP protein, including the Site-1 protease and Site-2 protease processing sites but not the N-terminal transcription factor domain, and acts genetically as a reduction-of-function mutation. These results

indicate that *daf-2*/InR signaling antagonizes ceSREBP function independent of ceSREBP proteolytic processing.

Two mutations, *sa680* and *sa709*, in the *pdk-1*/PDK1 (GI5353962 and GI5353964) gene which acts downstream of *daf-2*/InR (Paradis SM *et al.*, Genes & Dev. (1999) 13:1438-1452) also partially suppressed the pale intestine phenotype of ceSREBP feeding RNAi to a similar degree as the *daf-2 (e1370)* mutant. *pdk-1 (sa680)* and *pdk-1 (sa709)* are also dauer-constitutive mutations that are temperature-sensitive and were tested at the permissive temperature of 20°C for suppression of ceSREBP reduction of function in non-dauer animals.

The *daf-7*/TGF-β pathway acts in parallel with the *daf-2*/InR pathway in a non-redundant manner to inhibit dauer formation (Riddle DL and Albert PS, 1997,.in C. elegans *II* (eds. DL Riddle, T Blumenthal, BJ Meyer, and JR Priess), Cold Spring Harbor Laboratory, Cold Spring Harbor, NY, pp. 739-768.). Similar to *daf-2*/InR and *pdk-1*/PDK1 mutants, the *daf-7 (e1372)* temperature-sensitive mutation (GI1684866; Ren PC *et al.*, Science (1996) 274: 1389-1391) at the permissive temperatures 20°C and 15°C behaved as a weak suppressor of the pale intestine phenotype produced by RNAi of ceSREBP. The *daf-14*/Smad gene (GI6110600) acts downstream of *daf-7*/TGF-β, and a putative null mutation of *daf-14*, *m77* (Inoue T, and Thomas JH. Dev. Biol. (2000) 217: 192-204), partially suppressed the SREBP reduction of function. Interestingly, this suppression was nearly complete at 15°C.

We can conclude that in *C. elegans* the InR and TGF-β signaling pathways normally down-regulate the lipid biosynthesis and/or uptake activity that the ceSREBP pathway normally up-regulates.

WHAT IS CLAIMED IS:

1. An animal that is a fly or nematode that has been genetically modified to express or mis-express an SREBP pathway protein, or the progeny of said animal that has inherited said SREBP pathway protein expression or mis-expression.

2. The animal of Claim 1 that has been genetically modified by a method selected from the group consisting of transposon insertion mutagenesis, double-stranded RNA interference, and chemical mutagenesis.

3. The animal of Claim 1 wherein a heterologous promoter drives expression or mis-expression of said SREBP pathway protein.

4. The animal of Claim 3 wherein said promoter is selected from the group consisting of tissue-specific promoters, developmental-specific promoters, and inducible promoters.

5. The animal of Claim 4 wherein said animal is a fly and said promoter is selected from the group consisting of sevenless, eyeless, glass, dpp, heat shock, tTA-responsive, GAL4-responsive, and vestigal.

6. The animal of Claim 1 wherein said SREBP pathway protein is encoded by an SREBP pathway nucleic acid sequence linked to a nucleic acid sequence that encodes one or more selectable markers that allows detection of expression of said SREBP pathway protein.

7. The animal of Claim 1 wherein said expression or mis-expression of said SREBP pathway protein results in an identifiable phenotype.

8. The animal of Claim 1 wherein said SREBP pathway protein comprises an amino acid sequence selected from the group consisting of SEQ ID NOs:2, 4, 6, and 8, or a functionally-active fragment thereof.

9. The animal of Claim 8 wherein said SREBP pathway protein is encoded by part or all of a nucleic acid sequence selected from the group consisting of SEQ ID NOs:1, 3, 5, and 7.

10. The animal of Claim 7 wherein said nematode is heterozygous for deletion of SREBP.

11. The animal of Claim 7 wherein said animal is a nematode and said identifiable phenotype is a pale intestine phenotype or other intestinal defect.

12. A method for studying lipid metabolism comprising detecting the phenotype caused by the expression or mis-expression of said SREBP pathway protein in the animal of Claim 1.

13. The method of Claim 12 additionally comprising observing a second animal having the same genetic modification as the animal of Claim 1 which causes said expression or mis-expression of said SREBP pathway protein, and wherein said second animal additionally comprises a mutation in a gene of interest, wherein differences, if any, between the phenotype of the animal of Claim 1 and the phenotype of the second animal identifies the gene of interest as capable of modifying the function of the gene encoding said SREBP pathway protein.

14. The method of Claim 13 wherein said gene of interest is implicated in a pathway selected from the group consisting of cholesterol or fatty acid biosynthesis or metabolism, insulin signaling, and TGF-β signaling.

15. The method of Claim 13 wherein said animal is a nematode and wherein said phenotype is a pale intestine phenotype or other intestinal defect indicative of abnormalities in lipid biosynthesis or metabolism.

16. The method of Claim 13 wherein said animal is a nematode and wherein said method includes staining said nematode in vivo with a fluorescently-labelled fatty acid conjugate to measure lipid content within said nematode.

17. The method of Claim 16 wherein said fluorescently-labelled fatty acid conjugate is a BODIPY™-fatty acid conjugate.

18. The method of Claim 13 additionally comprising administering one or more compounds to said animal or its progeny and observing any changes in lipid content of said animal or its progeny.

19. A method for determining the lipid content of a living nematode comprising contacting said nematode with a BODIPY™ fatty acid conjugate to stain lipid and measuring fluorescence as an indication of lipid content.

20. The method of Claim 19 which is used in combination with a genetic screen for detection of mutations that affect lipid content.

21. The method of Claim 19 that additionally includes administering one or more compounds to said nematode or its progeny and observing any effect said compound has on lipid content.

22. An isolated nucleic acid molecule of less than 15 kb comprising a nucleic acid sequence selected from the group consisting of:

A) a nucleic acid sequence that encodes a polypeptide comprising at least 10 contiguous amino acids of the sequence of any one of SEQ ID NO:2, 4, and 6;

B) a nucleic acid sequence that encodes a polypeptide comprising at least 16 contiguous amino acids of SEQ ID NO:8; and

C) a nucleic acid sequence that encodes a polypeptide comprising at least 8 contiguous amino acids of residues 335 to 428 of SEQ ID NO:2.

23. The isolated nucleic acid molecule of Claim 22 that hybridizes under appropriate conditions to a nucleic acid sequence selected from the group consisting of SEQ ID NOs:1, 3, 5 and 7.

24. The isolated nucleic acid molecule of Claim 23 wherein said appropriate conditions comprise hybridization at 34°C in a buffer comprising 6X SSC / 0% formamide and a wash at 45°C in a buffer comprising 2X SSC.

25. A vector comprising the nucleic acid molecule of Claim 22.

26. A host cell comprising the vector of Claim 25.

27. The host cell of Claim 26 wherein said cell is a yeast cell.

28. A process for producing an SREBP pathway protein comprising culturing the host cell of Claim 26 under conditions suitable for expression of said SREBP pathway

protein and recovering said protein.

29. An isolated SREBP pathway protein produced by the process of Claim 28.

30. The isolated SREBP pathway protein of Claim 29 which is joined at its amino- or carboxy-terminus via a peptide bond to an amino acid sequence of a different protein.

31. A method of detecting a candidate molecule that binds to a polypeptide comprising SEQ ID 2, 4, 6, or 8 comprising:

(a) contacting said polypeptide with one or more candidate molecules under conditions conducive to binding; and

(b) detecting any binding that occurs between the candidate molecules and said polypeptide.

(12) INTERNATIONAL APPLICATION PUBLISHED UNDER THE PATENT COOPERATION TREATY (PCT)

(19) World Intellectual Property Organization
International Bureau

(43) International Publication Date
21 December 2000 (21.12.2000)

PCT

(10) International Publication Number
WO 00/76308 A1

(51) International Patent Classification[7]: A01K 67/00, 33/00, G01N 33/00

(21) International Application Number: PCT/US00/15880

(22) International Filing Date: 8 June 2000 (08.06.2000)

(25) Filing Language: English

(26) Publication Language: English

(30) Priority Data:

09/332,522	14 June 1999 (14.06.1999)	US
60/189,700	15 March 2000 (15.03.2000)	US

(71) Applicant *(for all designated States except US)*: **EXELIXIS, INC.** [US/US]; 280 East Grand Avenue, South San Francisco, CA 94080 (US).

(72) Inventors; and

(75) Inventors/Applicants *(for US only)*: **COSTA, Michael, A.** [US/US]; 86 Tioga Avenue, San Francisco, CA 94314 (US). **DOBERSTEIN, Stephen, Kohl** [US/US]; 2 Townsend Street, San Francisco, CA 94107 (US). **ELSON, Sarah** [US/US]; 1601 Florida Street, San Francisco, CA 94110 (US). **FERGUSON, Kimberly, Carr** [US/US]; 479 Ebken Street, Pacifica, CA 94044 (US). **HOMBURGER, Sheila, Akiko** [US/US]; 69 Glen Avenue, #304, Oakland, CA 94611 (US). **EBENS, Allen, James, Jr.** [US/US]; 1786 27th Avenue, San Francisco, CA 94122 (US). **KEEGAN, Kevin, Patrick** [US/US]; 1408 Sayre Street, San Leandro, CA 94579 (US). **STOUT, Thomas, J.** [US/US]; 128 Galewood Circle, San Francisco, CA 94131 (US).

(74) Agent: **BRUNELLE, Jan, P.**; Exelixis, Inc., 280 East Grand Avenue, South San Francisco, CA 94080 (US).

(81) Designated States *(national)*: AE, AL, AM, AT, AU, AZ, BA, BB, BG, BR, BY, CA, CH, CN, CR, CU, CZ, DE, DK, DM, EE, ES, FI, GB, GD, GE, GH, GM, HR, HU, ID, IL, IN, IS, JP, KE, KG, KP, KR, KZ, LC, LK, LR, LS, LT, LU, LV, MA, MD, MG, MK, MN, MW, MX, NO, NZ, PL, PT, RO, RU, SD, SE, SG, SI, SK, SL, TJ, TM, TR, TT, TZ, UA, UG, US, UZ, VN, YU, ZA, ZW.

(84) Designated States *(regional)*: ARIPO patent (GH, GM, KE, LS, MW, MZ, SD, SL, SZ, TZ, UG, ZW), Eurasian patent (AM, AZ, BY, KG, KZ, MD, RU, TJ, TM), European patent (AT, BE, CH, CY, DE, DK, ES, FI, FR, GB, GR, IE, IT, LU, MC, NL, PT, SE), OAPI patent (BF, BJ, CF, CG, CI, CM, GA, GN, GW, ML, MR, NE, SN, TD, TG).

Published:
— *With international search report.*

For two-letter codes and other abbreviations, refer to the "Guidance Notes on Codes and Abbreviations" appearing at the beginning of each regular issue of the PCT Gazette.

WO 00/76308 A1

(54) Title: ANIMAL MODELS AND METHODS FOR ANALYSIS OF LIPID METABOLISM AND SCREENING OF PHARMACEUTICAL AND PESTICIDAL AGENTS THAT MODULATE LIPID METABOLISM

(57) Abstract: Drosophila melanogaster and *C. elegans* that have been genetically modified to express or mis-express proteins involved in the sterol regulatory element binding protein (SREBP) pathway are described. These genetically modified animal models have identifiable phenotypes that make them useful in assays for studying lipid metabolism, other genes implicated in lipid metabolism, and compounds capable of modulating lipid metabolism pathways. Methods for studying lipid metabolism in living mematodes using fluorescently labeled fatty acid conjugates, such BODIPY™ fatty acid conjugates, are also described. Novel SREBP pathway nucleic acid and protein sequences are also described.

FIG.1B

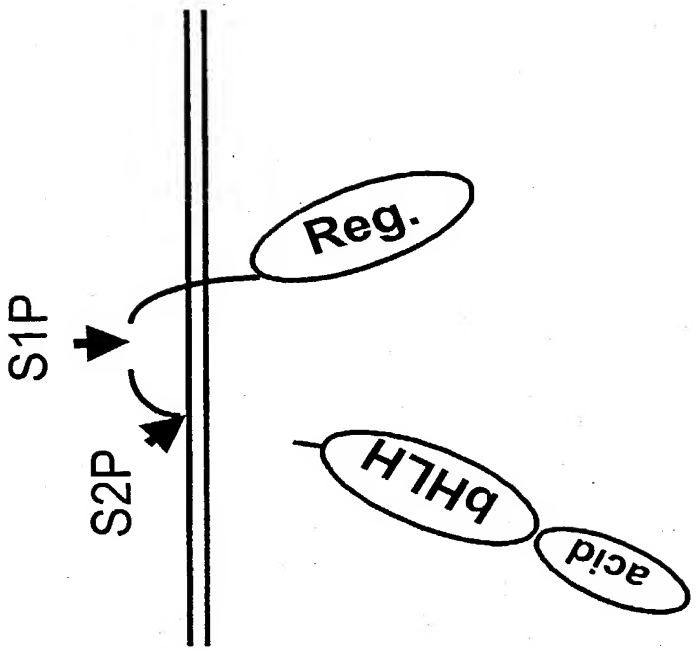
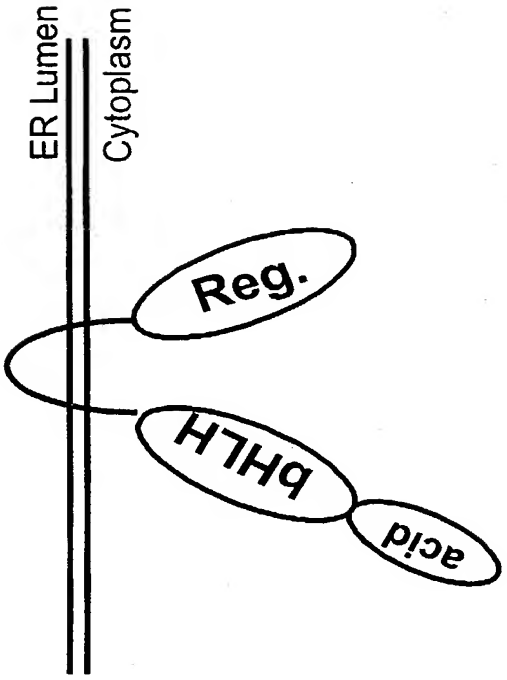


FIG.1A



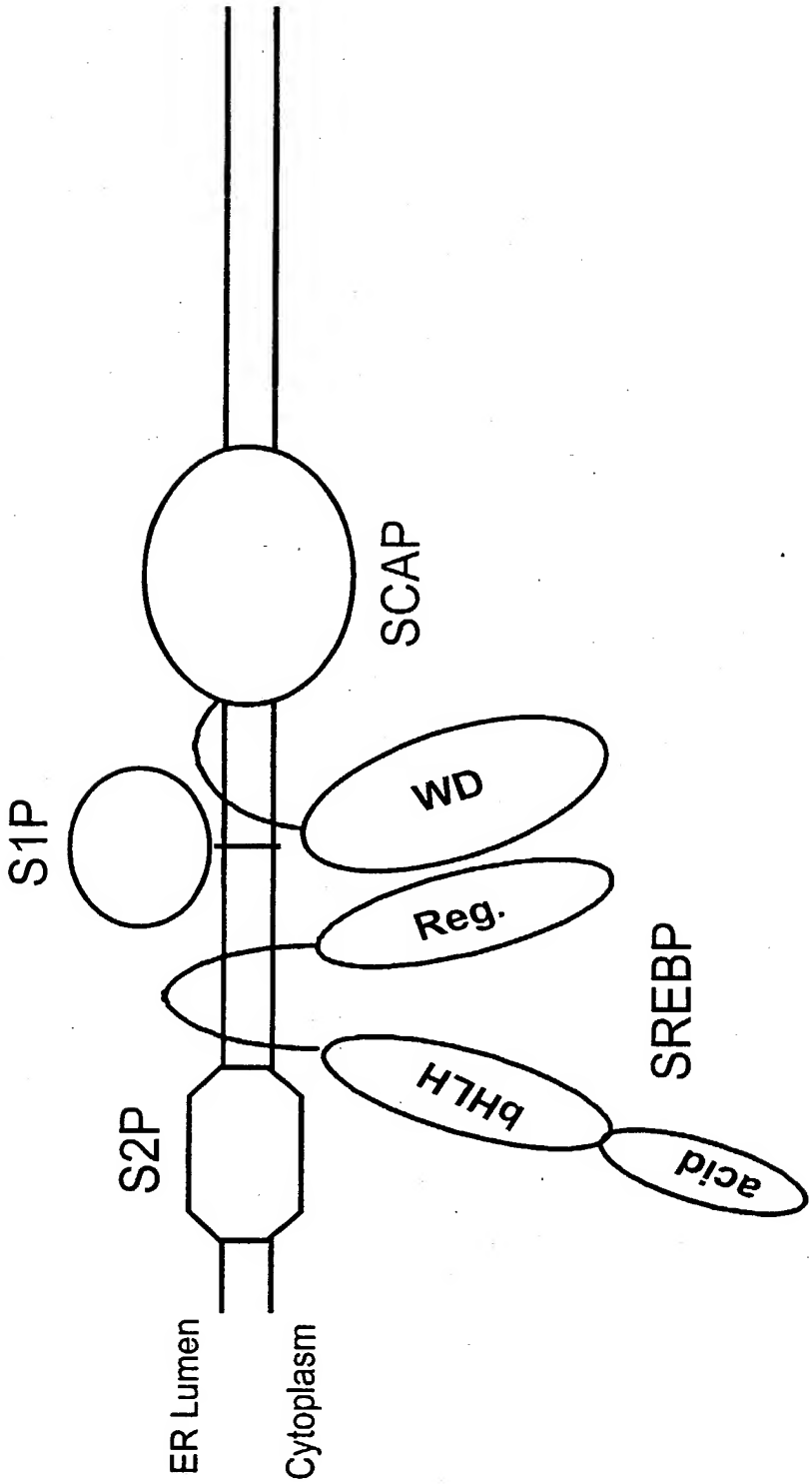


FIG.2

PTO/SB/01 (10-01)
Approved for use through 10/31/2002. OMB 0651-0032
U.S. Patent and Trademark Office; U.S. DEPARTMENT OF COMMERCE

DECLARATION FOR UTILITY OR DESIGN PATENT APPLICATION (37 CFR 1.63)

☒ Declaration Submitted With Initial Filing **OR** ☐ Declaration Submitted after Initial Filing (surcharge (37 CFR 1.16 (e)) required)

Attorney Docket Number	EX99-004C-US
First Named Inventor	Costa
COMPLETE IF KNOWN	
Application Number	/
Filing Date	
Group Art Unit	
Examiner Name	

As a below named inventor, I hereby declare that:

My residence, post office address, and citizenship are as stated below next to my name.

I believe I am the original and first inventor of the subject matter which is claimed and for which a patent is sought on the invention entitled:

Animal Models and Methods for Analysis of Lipid Metabolism and Screening of Pharmaceutical and Pesticidal Agents that Modulate Lipid Metabolism

(Title of the Invention)

the specification of which

☐ is attached hereto

OR

☒ was filed on (MM/DD/YYYY) 06/08/2000 as United States Application Number or PCT International Application Number PCT/US00/15880 and was amended on (MM/DD/YYYY) _____ (if applicable).

I hereby state that I have reviewed and understand the contents of the above identified specification, including the claims as amended specifically referred to above.

I acknowledge the duty to disclose information which is material to patentability as defined in 37 CFR 1.56, including for continuation-in-part applications, material information which became available between the filing date of the prior application and the national or PCT international filing date of the continuation-in-part application.

I hereby claim foreign priority benefits under 35 U.S.C. 119(a)-(d) or (f), or 365(b) of any foreign application(s) for patent, inventor's or plant breeder's rights certificate(s), or 365(a) of any PCT international application which designated at least one country other than the United States of America, listed below and have also identified below, by checking the box any foreign application(s) for patent, inventor's or plant breeder's rights certificate(s), or of any PCT international application having a filing date before that of the application on which priority is claimed.

Prior Foreign Application Number(s)	Country	Foreign Filing Date (MM/DD/YYYY) Country	Priority Not Claimed	Certified Copy Attached? YES	NO
			☐	☐	☐
			☐	☐	☐
			☐	☐	☐
			☐	☐	☐

☐ Additional foreign application numbers are listed on a supplemental priority data sheet PTO/SB/02B attached hereto:

PTO/SB/01 (10-01)
Approved for use through 10/31/2002. OMB 0651-0032
U.S. Patent and Trademark Office; U.S. DEPARTMENT OF COMMERCE

DECLARATION — Utility ~~or Design~~ Patent Application

Direct all correspondence to: ☒ Customer Number or Bar Code Label 23500 OR ☐ Correspondance address below

Name

Address

City | State | ZIP

Country | Telephone | Fax

I hereby declare that all statements made herein of my own knowledge are true and that all statements made on information and belief are believed to be true; and further that these statements were made with the knowledge that willful false statements and the like so made are punishable by fine or imprisonment, or both, under 18 U.S.C. 1001 and that such willful false statements may jeopardize the validity of the application or any patent issued thereon.

1-00

NAME OF SOLE OR FIRST INVENTOR: ☐ A petition has been filed for this unsigned inventor

Given Name (first and middle [if any]): Michael | Family Name or Surname: Costa

Inventor's Signature: Michael Costa | Date: 1/2/02

San Francisco | CA | US | US
Residence: City | State | Country | Citizenship

~~86 Tioga Avenue~~ 18 Hazelwood Ave.
Mailing Address

San Francisco | CA | ~~94134~~ 94112 | US
City | State | Zip | Country

NAME OF SECOND INVENTOR: ☐ A petition has been filed for this unsigned inventor

Given Name (first and middle [if any]): Stephen Kohl | Family Name or Surname: Doberstein

Inventor's Signature | Date

South San Francisco | CA | US | US
Residence: City | State | Country | Citizenship

827 Linden Avenue
Mailing Address

South San Francisco | CA | 94080 | US
City | State | Zip | Country

☒ Additional inventors are being named on the 2 supplemental Additional Inventor(s) sheet(s) PTO/SB/02A attached hereto.

PTO/SB/01 (10-01)
Approved for use through 10/31/2002. OMB 0651-0032
U.S. Patent and Trademark Office; U.S. DEPARTMENT OF COMMERCE

DECLARATION — Utility or Design Patent Application

Direct all correspondence to: ☒ Customer Number or Bar Code Label 23500 *OR* ☐ Correspondance address below

Name

Address

City	**State**	**ZIP**

Country	**Telephone**	**Fax**

I hereby declare that all statements made herein of my own knowledge are true and that all statements made on information and belief are believed to be true; and further that these statements were made with the knowledge that willful false statements and the like so made are punishable by fine or imprisonment, or both, under 18 U.S.C. 1001 and that such willful false statements may jeopardize the validity of the application or any patent issued thereon.

NAME OF SOLE OR FIRST INVENTOR: ☐ A petition has been filed for this unsigned inventor

Given Name (first and middle [if any]) Michael	**Family Name or Surname** Costa

Inventor's Signature	**Date**

San Francisco **Residence: City**	CA **State**	US **Country**	US **Citizenship**

86 Tioga Avenue
Mailing Address

San Francisco **City**	CA **State**	94134 **Zip**	US **Country**

NAME OF SECOND INVENTOR: ☐ A petition has been filed for this unsigned inventor

Given Name (first and middle [if any]) Stephen Kohl	**Family Name or Surname** Doberstein

Inventor's Signature S Kohl	**Date** 1/8/02

South San Francisco **Residence: City** CA.	CA **State**	US **Country**	US **Citizenship**

827 Linden Avenue
Mailing Address

South San Francisco **City**	CA **State**	94080 **Zip**	US **Country**

☒ **Additional inventors are being named on the 2** supplemental Additional Inventor(s) sheet(s) PTO/SB/02A attached hereto.

2-00

PTO/SB/02A (11-00)

Please type a plus sign (+) inside this box → +

Approved for use through 10/31/2002. OMB 0651-0032
U.S. Patent and Trademark Office; U.S. DEPARTMENT OF COMMERCE

DECLARATION

ADDITIONAL INVENTOR(S)
Supplemental Sheet
Page 1 of 2

3-00

Name of Additional Joint Inventor, if any:	☐ A petition has been filed for this unsigned inventor		
Given Name (first and middle [if any])	Family Name or Surname		
Sarah	Elson		
Inventor's Signature			**Date** 1/24/02
Residence: City San Francisco	**State** CA	**Country** US	**Citizenship** US
Mailing Address 1604 Florida Street CA.			
Mailing Address			
City San Francisco	**State** CA	**ZIP** 94110	**Country** US

4-00

Name of Additional Joint Inventor, if any:	☐ A petition has been filed for this unsigned inventor		
Given Name (first and middle [if any])	Family Name or Surname		
Kimberly Carr	Ferguson		
Inventor's Signature Kimberly Carr Ferguson			**Date** 1/4/02
Residence: City ~~Pacifica~~ SAN FRANCISCO	**State** CA CA	**Country** US	**Citizenship** US
Mailing Address ~~479 Ebken Street~~ 2018 10TH AVENUE			
Mailing Address			
City SAN FRANCISCO ~~Pacifica~~	**State** CA	**Zip** ~~94044~~ 94116	**Country** US

Name of Additional Joint Inventor, if any:	☐ A petition has been filed for this unsigned inventor		
Given Name (first and middle [if any])	Family Name or Surname		
Sheila Akiko	Homburger		
Inventor's Signature			**Date**
Residence: City Oakland	**State** CA	**Country** US	**Citizenship** US
Mailing Address 69 Glen Avenue			
Mailing Address			
City Oakland	**State** CA	**Zip** 94611	**Country** US

PTO/SB/02A (11-00)

Please type a plus sign (+) inside this box → [+]

Approved for use through 10/31/2002. OMB 0651-0032
U.S. Patent and Trademark Office; U.S. DEPARTMENT OF COMMERCE

DECLARATION

ADDITIONAL INVENTOR(S)
Supplemental Sheet
Page 1 of 2

Name of Additional Joint Inventor, if any:	☐ A petition has been filed for this unsigned inventor		
Given Name (first and middle [if any])		Family Name or Surname	
Sarah		Elson	
Inventor's Signature			**Date**
Residence: City San Francisco	**State** CA	**Country** US	**Citizenship** US
Mailing Address 1604 Florida Street			
Mailing Address			
City San Francisco	**State** CA	**ZIP** 94110	**Country** US

Name of Additional Joint Inventor, if any:	☐ A petition has been filed for this unsigned inventor		
Given Name (first and middle [if any])		Family Name or Surname	
Kimberly Carr		Ferguson	
Inventor's Signature			**Date**
Residence: City Pacifica	**State** CA	**Country** US	**Citizenship** US
Mailing Address 479 Ebken Street			
Mailing Address			
City Pacifica	**State** CA	**Zip** 94044	**Country** US

Name of Additional Joint Inventor, if any:	☐ A petition has been filed for this unsigned inventor		
Given Name (first and middle [if any])		Family Name or Surname	
Sheila ~~Akiko~~ Homburger		~~Homburger~~ Avery	
Inventor's Signature Sheila Homburger Avery			**Date** 1/9/2002
Residence: City Oakland	**State** CA	**Country** US	**Citizenship** US
Mailing Address 69 Glen Avenue CA			
Mailing Address			
City Oakland	**State** CA	**Zip** 94611	**Country** US

5.0

PTO/SB/02A (11-00)

Please type a plus sign (+) inside this box → [+]

Approved for use through 10/31/2002. OMB 0651-0032
U.S. Patent and Trademark Office; U.S. DEPARTMENT OF COMMERCE
Under the Paperwork Reduction Act of 1995, no persons are required to respond to a collection of information unless it contains a valid OMB control number.

DECLARATION

ADDITIONAL INVENTOR(S)
Supplemental Sheet
Page 2 of 2

Name of Additional Joint Inventor, if any:	☐ A petition has been filed for this unsigned inventor		
Given Name (first and middle [if any])	Family Name or Surname		
Allen James Jr.	Ebens		
Inventor's Signature Allen James Ebens Jr.			**Date** 01/04/02
Residence: City San Francisco	**State** CA	**Country** US	**Citizenship** US
Mailing Address 1786 27th Avenue CA.			
Mailing Address			
City San Francisco	**State** CA	**ZIP** 94122	**Country** US

Name of Additional Joint Inventor, if any:	☐ A petition has been filed for this unsigned inventor		
Given Name (first and middle [if any])	Family Name or Surname		
Kevin Patrick	Keegan		
Inventor's Signature			**Date**
Residence: City San Diego	**State** CA	**Country** US	**Citizenship** US
Mailing Address 8989 Renato Street			
Mailing Address			
City San Diego	**State** CA	**Zip** 92129	**Country** US

Name of Additional Joint Inventor, if any:	☐ A petition has been filed for this unsigned inventor		
Given Name (first and middle [if any])	Family Name or Surname		
Thomas J.	Stout		
Inventor's Signature Thomas J Stout			**Date** 01/05/02
Residence: City San Francisco	**State** CA	**Country** US	**Citizenship** US
Mailing Address 128 Galewood Circle CA.			
Mailing Address			
City San Francisco	**State** CA	**Zip** 94131	**Country** US

Burden Hour Statement: This form is estimated to take 21 minutes to complete. Time will vary depending upon the needs of the individual case. Any comments on the amount of time you are required to complete this form should be sent to the Chief Information Officer, U.S. Patent and Trademark Office, Washington, DC 20231. DO NOT SEND FEES OR COMPLETED FORMS TO THIS ADDRESS. SEND TO: Assistant Commissioner for Patents, Washington, DC 20231.

PTO/SB/02A (11-00)

Please type a plus sign (+) inside this box → [+]

Approved for use through 10/31/2002. OMB 0651-0032
U.S. Patent and Trademark Office; U.S. DEPARTMENT OF COMMERCE
Under the Paperwork Reduction Act of 1995, no persons are required to respond to a collection of information unless it contains a valid OMB control number.

DECLARATION

ADDITIONAL INVENTOR(S)
Supplemental Sheet
Page 2 of 2

Name of Additional Joint Inventor, if any:	☐ A petition has been filed for this unsigned inventor		
Given Name (first and middle [if any])	Family Name or Surname		
Allen James Jr.	Ebens		
Inventor's Signature			**Date**
Residence: City San Francisco	**State** CA	**Country** US	**Citizenship** US
Mailing Address 1786 27th Avenue			
Mailing Address			
City San Francisco	**State** CA	**ZIP** 94122	**Country** US

Name of Additional Joint Inventor, if any:	☐ A petition has been filed for this unsigned inventor		
Given Name (first and middle [if any])	Family Name or Surname		
Kevin Patrick	Keegan		
Inventor's Signature K. P. Keegan			**Date** 1/4/02
Residence: City San Diego	**State** CA	**Country** US	**Citizenship** US
Mailing Address 8989 Renato Street CT			
Mailing Address			
City San Diego	**State** CA	**Zip** 92129	**Country** US

Name of Additional Joint Inventor, if any:	☐ A petition has been filed for this unsigned inventor		
Given Name (first and middle [if any])	Family Name or Surname		
Thomas J.	Stout		
Inventor's Signature			**Date**
Residence: City San Francisco	**State** CA	**Country** US	**Citizenship** US
Mailing Address 128 Galewood Circle			
Mailing Address			
City San Francisco	**State** CA	**Zip** 94131	**Country** US

Burden Hour Statement: This form is estimated to take 21 minutes to complete. Time will vary depending upon the needs of the individual case. Any comments on the amount of time you are required to complete this form should be sent to the Chief Information Officer, U.S. Patent and Trademark Office, Washington, DC 20231. DO NOT SEND FEES OR COMPLETED FORMS TO THIS ADDRESS. SEND TO: Assistant Commissioner for Patents, Washington, DC 20231.

SEQUENCE LISTING

<110> EXELIXIS, INC.

<120> ANIMAL MODELS AND METHODS FOR ANALYSIS OF LIPID METABOLISM AND SCREENING OF PHARMACEUTICAL AND PESTICIDAL AGENTS THAT MODULATE LIPID METABOLISM

<130> SREBP-INT

<140> EX99-004C-PCT
<141> 2000-06-08

<150> 09/332,522
<151> 1999-06-14

<150> 60/189,700
<151> 2000-03-15

<160> 8

<170> PatentIn Ver. 2.1

<210> 1
<211> 3419
<212> DNA
<213> Caenorhabditis elegans

<400> 1
ggtttaatta cccaagtttg agaatgaacg aagaattcga gggagacgtc cctatgtcgg 60
atccgtttct ctcattggtc acaaaattgg atgatattgc gccatttcca aataacgacc 120
cgctcgattt tgacatggag cacaactggc aagagcccgg accatcacaa caaccggatc 180
catcaattcc cggaaatcaa cacagtccgc cacaggaata ttatgatatt gatggtcaac 240
gagacgtaag caccttacac tccctgctca accacaacaa cgacgacttc ttctcaatgc 300
gattttcccc gccaaacttt gatctcggcg gaggccgtgg accttctcta gccgccaccc 360
aacaattatc tggagaaggt cctgcaagta tgcttaaccc cttacaaaca tctccaccaa 420
gtggaggtta ccccccggca gatgcctaca gacctctatc acttgctcaa caactcgccg 480
cgccagcgat gactccacat caggcagcgt cgcttttgt taatactaat ggaattgatc 540
aaaagaattt cactcatgca atgctatctt caccacacca tacctcaatg acttctcaac 600
catatacaga agccatggga catatcaacg ggtacatgtc tccatacgac caagctcaag 660
gcccatcagg accatcatat tactcacaac accatcaatc tccaccacct catcaccacc 720
atcaccaccc gatgccaaaa atccatgaga accctgaaca agtggcatct ccatcgattg 780
aagatgctcc agagacgaaa ccaactcatt tggttgaacc acaaagtcca aaaagcccgc 840
agaatatgaa agaggagctt cttcggttac tagttaacat gtctccgagt gaagttgaac 900
ggttaaagaa taaaaaatca ggagcatgtt cagcgacgaa tgggccatcg aggagtaagg 960
agaaggcggc gaagattgtg attcaggaga cagcggaagg ggatgaagat gaggatgatg 1020
aggatagtga ttccggggag actatgtctc agggaactac tattattgtt cgaagaccaa 1080
aaaccgagcg tcgtacggca cacaatctca tcgaaaagaa gtatagatgc tcaataaatg 1140

atcgaattca acagctgaaa gtacttttgt gtggggatga agctaagctt tcaaaatcgg 1200
caacactacg acgggctatt gaacatatcg aggaggttga acacgagaat caggtgttga 1260
agcatcatgt tgaacaaatg agaaagacac tgcagaataa tcgattaccg tacccggaac 1320
caattcaata cactgaatac tctgcccgat cacccgtcga atcatctcct tctccaccta 1380
gaaatgagag aaaacgatca cgaatgagca caacgactcc tatgaagaat ggaactagag 1440
atggatcttc gaaagttacc ctttttgcga tgctcctagc agttctgatt tttaatccga 1500
ttggattgct cgctggaagt gcgatattct caaaagccgc tgcagaagct ccgattgcct 1560
ccccgttcga gcatggaaga gtgattgatg acccggatgg aactagcact cggacgcttt 1620
tctgggaagg gagtatcatc aatatgagct atgtctgggt gttcaacatc ttaatgatca 1680
tatatgtggt tgtcaaactg ctgatccatg gtgaccctgt tcaagacttc atgtccgttt 1740
catggcagac ttttgtgacg actcgagaga aggcgagagc cgagttgaac tctggaaatt 1800
tgaaagatgc tcagagaaag ttctgcgagt gtcttgcaac gttggatcga tcgcttccat 1860
caccgggggt tgattcggtg ttttcggttg gctgggaatg cgttcgacat cttttgaatt 1920
ggttgtggat cgggagatac atcgcaagaa ggcgcaggtc caccacgaag cctgtctcag 1980
tcgtttgtag gagtcatgcg cagactgcag ttctctatca tgaaattcat cagctccatc 2040
taatgggtat cactggaaac ttcgaagaca cctatgaacc atccgcccta acgggcctct 2100
tcatgtccct ctgtgcagta aaccttgctg aagctgccgg agcatcaaac gacggacttc 2160
cacgcgccgt catggctcag atctacattt ctgcatccat ccaatgccgt ttggctcttc 2220
cgaacctact cgcaccattc ttctcgggat actttttacg aagagctcga aggcacgtgc 2280
gtcgagctcc ggagcactcg gtgtcccatt tgttatggat cttccatcca gcgacaagaa 2340
agttcatgtc agatgcgaaa aggttggagc atgtgttgag ctcgaagcag aagcagttga 2400
gatttgggtc ttttgtggaa gatgagcaat tatccccact tgctcgaatc cgaacaacgc 2460
tgaaagtgta cctactctcc aaacttgtac aggaacttgt cggtggtgac gagatcttta 2520
caaaaaatgt ggaacgcatc ctaaatgaca atgaccgtct cgatgatgaa gtagacgtgg 2580
ttgatgtttc aagacttttg gtgacaattt caacgcagtg cgctgccatt ttgactaatg 2640
agaaggatga gtcagcgaaa ttcggaacct ggatctctcg aaacggagat gcttgttgca 2700
catggtggac gcacgttctg acatgtggaa tctattggag gagtaacaag aatgagctgg 2760
cacggcaaca ctattcactg atcaggaact gtccgccgaa gattttgaca gacaatctgg 2820
gtttggcggt tggccacgcg ttgtgtgctc gcaagatttg catagatgac cgagattccc 2880
cgaaagtcag tcaatacgtg tgcattcaca caaagaagtc gctcgaatcc ctccgactat 2940
tctccacatc atcgcgagca tcaggtgtgg tgtctggaat tcaggaaggt acacgccgaa 3000
tggcctacga atggattatg aactcgctgc tcgacgcgtg gcgttccaat ctattcgcat 3060
cgaaacccta ctggacacaa agcttcaagg gacaatccac gtttagtacg ctttatcaag 3120
aggcgtataa tcattatgcg attattaatg ggacaagggg agattgttgg agactatttg 3180
tctacgagct cacgtgccga atgctcaacg gagccaaccc acaagccacg tggtcaggcg 3240
yccgacgcgt tcgatctaca aaaatggacg cggtccgagg aagagtgagc atgcgacgct 3300
cggctcaacc ggacgcattt catcttcata cactggttaa actacatact tctatggatc 3360
tttgaattga acaaaaaatg attttattca gaataatgat aaatacgatt atatataaa 3419

<210> 2
<211> 1113
<212> PRT
<213> Caenorhabditis elegans

<400> 2
Met Asn Glu Glu Phe Glu Gly Asp Val Pro Met Ser Asp Pro Phe Leu
1 5 10 15

Ser Leu Val Thr Lys Leu Asp Asp Ile Ala Pro Phe Pro Asn Asn Asp
20 25 30

Pro Leu Asp Phe Asp Met Glu His Asn Trp Gln Glu Pro Gly Pro Ser
35 40 45

Gln Gln Pro Asp Pro Ser Ile Pro Gly Asn Gln His Ser Pro Pro Gln
50 55 60

Glu Tyr Tyr Asp Ile Asp Gly Gln Arg Asp Val Ser Thr Leu His Ser
65 70 75 80

Leu Leu Asn His Asn Asn Asp Asp Phe Phe Ser Met Arg Phe Ser Pro
85 90 95

Pro Asn Phe Asp Leu Gly Gly Gly Arg Gly Pro Ser Leu Ala Ala Thr
100 105 110

Gln Gln Leu Ser Gly Glu Gly Pro Ala Ser Met Leu Asn Pro Leu Gln
115 120 125

Thr Ser Pro Pro Ser Gly Gly Tyr Pro Pro Ala Asp Ala Tyr Arg Pro
130 135 140

Leu Ser Leu Ala Gln Gln Leu Ala Ala Pro Ala Met Thr Pro His Gln
145 150 155 160

Ala Ala Ser Leu Phe Val Asn Thr Asn Gly Ile Asp Gln Lys Asn Phe
165 170 175

Thr His Ala Met Leu Ser Ser Pro His His Thr Ser Met Thr Ser Gln
180 185 190

Pro Tyr Thr Glu Ala Met Gly His Ile Asn Gly Tyr Met Ser Pro Tyr
195 200 205

Asp Gln Ala Gln Gly Pro Ser Gly Pro Ser Tyr Tyr Ser Gln His His
210 215 220

Gln Ser Pro Pro Pro His His His His His His Pro Met Pro Lys Ile
225 230 235 240

His Glu Asn Pro Glu Gln Val Ala Ser Pro Ser Ile Glu Asp Ala Pro
245 250 255

Glu Thr Lys Pro Thr His Leu Val Glu Pro Gln Ser Pro Lys Ser Pro
260 265 270

Gln Asn Met Lys Glu Glu Leu Leu Arg Leu Leu Val Asn Met Ser Pro
275 280 285

Ser Glu Val Glu Arg Leu Lys Asn Lys Lys Ser Gly Ala Cys Ser Ala
290 295 300

Thr Asn Gly Pro Ser Arg Ser Lys Glu Lys Ala Ala Lys Ile Val Ile
305 310 315 320

Gln Glu Thr Ala Glu Gly Asp Glu Asp Glu Asp Asp Glu Asp Ser Asp
325 330 335

Ser Gly Glu Thr Met Ser Gln Gly Thr Thr Ile Ile Val Arg Arg Pro
340 345 350

Lys Thr Glu Arg Arg Thr Ala His Asn Leu Ile Glu Lys Lys Tyr Arg
355 360 365

Cys Ser Ile Asn Asp Arg Ile Gln Gln Leu Lys Val Leu Leu Cys Gly
370 375 380

Asp Glu Ala Lys Leu Ser Lys Ser Ala Thr Leu Arg Arg Ala Ile Glu
385 390 395 400

His Ile Glu Glu Val Glu His Glu Asn Gln Val Leu Lys His His Val
405 410 415

Glu Gln Met Arg Lys Thr Leu Gln Asn Asn Arg Leu Pro Tyr Pro Glu
420 425 430

Pro Ile Gln Tyr Thr Glu Tyr Ser Ala Arg Ser Pro Val Glu Ser Ser
435 440 445

Pro Ser Pro Pro Arg Asn Glu Arg Lys Arg Ser Arg Met Ser Thr Thr
450 455 460

Thr Pro Met Lys Asn Gly Thr Arg Asp Gly Ser Ser Lys Val Thr Leu
465 470 475 480

Phe Ala Met Leu Leu Ala Val Leu Ile Phe Asn Pro Ile Gly Leu Leu
485 490 495

Ala Gly Ser Ala Ile Phe Ser Lys Ala Ala Ala Glu Ala Pro Ile Ala
500 505 510

Ser Pro Phe Glu His Gly Arg Val Ile Asp Asp Pro Asp Gly Thr Ser
515 520 525

Thr Arg Thr Leu Phe Trp Glu Gly Ser Ile Ile Asn Met Ser Tyr Val
530 535 540

Trp Val Phe Asn Ile Leu Met Ile Ile Tyr Val Val Val Lys Leu Leu
545 550 555 560

Ile His Gly Asp Pro Val Gln Asp Phe Met Ser Val Ser Trp Gln Thr
565 570 575

Phe Val Thr Thr Arg Glu Lys Ala Arg Ala Glu Leu Asn Ser Gly Asn
580 585 590

Leu Lys Asp Ala Gln Arg Lys Phe Cys Glu Cys Leu Ala Thr Leu Asp
595 600 605

Arg Ser Leu Pro Ser Pro Gly Val Asp Ser Val Phe Ser Val Gly Trp
610 615 620

Glu Cys Val Arg His Leu Leu Asn Trp Leu Trp Ile Gly Arg Tyr Ile
625 630 635 640

Ala Arg Arg Arg Arg Ser Thr Thr Lys Pro Val Ser Val Val Cys Arg
645 650 655

Ser His Ala Gln Thr Ala Val Leu Tyr His Glu Ile His Gln Leu His
660 665 670

Leu Met Gly Ile Thr Gly Asn Phe Glu Asp Thr Tyr Glu Pro Ser Ala
675 680 685

Leu Thr Gly Leu Phe Met Ser Leu Cys Ala Val Asn Leu Ala Glu Ala
690 695 700

Ala Gly Ala Ser Asn Asp Gly Leu Pro Arg Ala Val Met Ala Gln Ile
705 710 715 720

Tyr Ile Ser Ala Ser Ile Gln Cys Arg Leu Ala Leu Pro Asn Leu Leu
725 730 735

Ala Pro Phe Phe Ser Gly Tyr Phe Leu Arg Arg Ala Arg Arg His Val
740 745 750

Arg Arg Ala Pro Glu His Ser Val Ser His Leu Leu Trp Ile Phe His
755 760 765

Pro Ala Thr Arg Lys Phe Met Ser Asp Ala Lys Arg Leu Glu His Val
770 775 780

Leu Ser Ser Lys Gln Lys Gln Leu Arg Phe Gly Ser Phe Val Glu Asp
785 790 795 800

Glu Gln Leu Ser Pro Leu Ala Arg Ile Arg Thr Thr Leu Lys Val Tyr
805 810 815

Leu Leu Ser Lys Leu Val Gln Glu Leu Val Gly Gly Asp Glu Ile Phe
820 825 830

Thr Lys Asn Val Glu Arg Ile Leu Asn Asp Asn Asp Arg Leu Asp Asp
835 840 845

Glu Val Asp Val Val Asp Val Ser Arg Leu Leu Val Thr Ile Ser Thr
850 855 860

Gln Cys Ala Ala Ile Leu Thr Asn Glu Lys Asp Glu Ser Ala Lys Phe
865 870 875 880

Gly Thr Trp Ile Ser Arg Asn Gly Asp Ala Cys Cys Thr Trp Trp Thr
885 890 895

His Val Leu Thr Cys Gly Ile Tyr Trp Arg Ser Asn Lys Asn Glu Leu
900 905 910

Ala Arg Gln His Tyr Ser Leu Ile Arg Asn Cys Pro Pro Lys Ile Leu
915 920 925

Thr Asp Asn Leu Gly Leu Ala Val Gly His Ala Leu Cys Ala Arg Lys
930 935 940

Ile Cys Ile Asp Asp Arg Asp Ser Pro Lys Val Ser Gln Tyr Val Cys
945 950 955 960

Ile His Thr Lys Lys Ser Leu Glu Ser Leu Arg Leu Phe Ser Thr Ser
965 970 975

Ser Arg Ala Ser Gly Val Val Ser Gly Ile Gln Glu Gly Thr Arg Arg
980 985 990

Met Ala Tyr Glu Trp Ile Met Asn Ser Leu Leu Asp Ala Trp Arg Ser
995 1000 1005

Asn Leu Phe Ala Ser Lys Pro Tyr Trp Thr Gln Ser Phe Lys Gly Gln
1010 1015 1020

Ser Thr Phe Ser Thr Leu Tyr Gln Glu Ala Tyr Asn His Tyr Ala Ile
1025 1030 1035 1040

Ile Asn Gly Thr Arg Gly Asp Cys Trp Arg Leu Phe Val Tyr Glu Leu
1045 1050 1055

Thr Cys Arg Met Leu Asn Gly Ala Asn Pro Gln Ala Thr Trp Ser Gly
1060 1065 1070

Xaa Arg Arg Val Arg Ser Thr Lys Met Asp Ala Val Arg Gly Arg Val
1075 1080 1085

Ser Met Arg Arg Ser Ala Gln Pro Asp Ala Phe His Leu His Thr Leu
1090 1095 1100

Val Lys Leu His Thr Ser Met Asp Leu
1105 1110

<210> 3
<211> 1971
<212> DNA
<213> Drosophila melanogaster

<400> 3
cggcacgagg attaatgctg atttctggtc tggactacac agcattgctg gtataaggag 60
tcgggaccag aggagtaaga tttcgggaag gaatcccgtc cggtagggac tactagcatt 120
cgcaagtgac gtccagcaac cggaggaccc ccaactgtag aatccgcatc accatcctaa 180
tcccaacaaa ccaatgacat cttgagacct caccagccat ggatcccttc gtgttcttca 240
tagtactggc atcgctttat ggcgttcttt actttttcga ccgcttcttc aagagttgca 300
tgcactaccc gtacgatgcc ttcctcaaga acaccgggct gagtataaat ttcatgagcc 360
tccactggca cacgagtgcc tttaacagga ccctcctacg ctggggatct gccggtaaca 420
gctgcacccg gagagtaatg atcaccagct ttaatgtagg agtcctggtc acctttttctc 480
tgctcccgat cggtctgatc ctgctcattg ccactatctt cagcagtggt gaacaagaca 540
gctcttcgtc tgtatcctcg cccgttggag tccctgtgca gctggaaatt ctactgcccg 600
gcgtcaactt gccgttggag gagatcggat actacatcac aacccttgtg ctctgcttgg 660
tggtgcacga gatgggacac gccctggccg ctgtgatgga ggatgtgcct gtcaccgggt 720
ttggaataaa gttcatcttc tgcctgccgt tagcatacac ggagctctcc cacgaccact 780
taaacagtct acgttggttc cgcaagctac gtgttctgtg cgctggaatc tggcataatt 840
ttgtgttcgc tggcgtgtgc tatctcttaa tctcaacggt gggaatcact atgtcacctt 900
tgtacgctta caaccaacac gtagtggtca ctgaactaac aaggaaatcc ccgctgaggg 960
gagagcgcgg cttgcaagtg gacaatcaaa taacccaagt aaacggctgc ccagtaaaca 1020
gcgaggagag ttgggtgaca tgcctgcaga actctctgaa gctcaagccg ggctactgtg 1080
tgagtgcgga cttcgtgcag cttaacgacg aaagcagcgc catctcacat catagcattg 1140
atggtcagct acagtgctgt gatgaactaa atccgaacgt aagctgcttc gaggtggtgg 1200
aggacgcaaa tggagatgtg ccggtggagc tgccgcagca tgtatgtctc aatgtgcgcc 1260
gcactttgga ggaggtctcc gagcactgct cgtccggagt ttgcaacgag ggattctgcc 1320
tacgaccgct tatacgaaat atcactgcca taatgacgtt caagcgacag aattttcgcg 1380
gagagaagct gccgccggtg atctatgtgg gccatccatg ggatgtcact cgaactgtgg 1440
aggtatccgc ctttgtgccg agatatagct tattaaaggc agcctggccg gatgcctggc 1500

tgctgctcct caagtataac gtggtcttca gcataggatt ggcgttgatc aatgccattc 1560
cctgctttgg tttcgatggc gcccacatta ccagcaccgt gatacacagc ttcttggtgg 1620
gcagagtgga tcagcatgcc aagagagata tcatctcgtt gataatcacc agcgtgggtt 1680
ccctttctctt tgcactggcc ctgcttaagg tggcctggtt gagttttctg cgacccctgc 1740
tttaagaact gaaatggaaa actgaaatgg atcctgggag ttcaactccc tgcaaagacg 1800
ctagactgct atttcacctt cacgaaacac acaaaaacac agcgaattgt agcacctcaa 1860
agattcgata gctttttgtc atagtcctta gtcttaactc gtatttattt tcgtacggtt 1920
gtcgagctca aaaataaaat caaattaagc taaaaaaaaa aaaaaaaaaa c 1971

<210> 4
<211> 508
<212> PRT
<213> Drosophila melanogaster

<400> 4
Met Asp Pro Phe Val Phe Phe Ile Val Leu Ala Ser Leu Tyr Gly Val
1 5 10 15

Leu Tyr Phe Phe Asp Arg Phe Phe Lys Ser Cys Met His Tyr Pro Tyr
20 25 30

Asp Ala Phe Leu Lys Asn Thr Gly Leu Ser Ile Asn Phe Met Ser Leu
35 40 45

His Trp His Thr Ser Ala Phe Asn Arg Thr Leu Leu Arg Trp Gly Ser
50 55 60

Ala Gly Asn Ser Cys Thr Arg Arg Val Met Ile Thr Ser Phe Asn Val
65 70 75 80

Gly Val Leu Val Thr Phe Ser Leu Leu Pro Ile Gly Leu Ile Leu Leu
85 90 95

Ile Ala Thr Ile Phe Ser Ser Gly Glu Gln Asp Ser Ser Ser Ser Val
100 105 110

Ser Ser Pro Val Gly Val Pro Val Gln Leu Glu Ile Leu Leu Pro Gly
115 120 125

Val Asn Leu Pro Leu Glu Glu Ile Gly Tyr Tyr Ile Thr Thr Leu Val
130 135 140

Leu Cys Leu Val Val His Glu Met Gly His Ala Leu Ala Ala Val Met
145 150 155 160

Glu Asp Val Pro Val Thr Gly Phe Gly Ile Lys Phe Ile Phe Cys Leu
165 170 175

Pro Leu Ala Tyr Thr Glu Leu Ser His Asp His Leu Asn Ser Leu Arg
180 185 190

Trp Phe Arg Lys Leu Arg Val Leu Cys Ala Gly Ile Trp His Asn Phe
195 200 205

Val Phe Ala Gly Val Cys Tyr Leu Leu Ile Ser Thr Val Gly Ile Thr
210 215 220

Met Ser Pro Leu Tyr Ala Tyr Asn Gln His Val Val Val Thr Glu Leu
225 230 235 240

Thr Arg Lys Ser Pro Leu Arg Gly Glu Arg Gly Leu Gln Val Asp Asn
245 250 255

Gln Ile Thr Gln Val Asn Gly Cys Pro Val Asn Ser Glu Glu Ser Trp
260 265 270

Val Thr Cys Leu Gln Asn Ser Leu Lys Leu Lys Pro Gly Tyr Cys Val
275 280 285

Ser Ala Asp Phe Val Gln Leu Asn Asp Glu Ser Ser Ala Ile Ser His
290 295 300

His Ser Ile Asp Gly Gln Leu Gln Cys Cys Asp Glu Leu Asn Pro Asn
305 310 315 320

Val Ser Cys Phe Glu Val Val Glu Asp Ala Asn Gly Asp Val Pro Val
325 330 335

Glu Leu Pro Gln His Val Cys Leu Asn Val Arg Arg Thr Leu Glu Glu
340 345 350

Val Ser Glu His Cys Ser Ser Gly Val Cys Asn Glu Gly Phe Cys Leu
355 360 365

Arg Pro Leu Ile Arg Asn Ile Thr Ala Ile Met Thr Phe Lys Arg Gln
370 375 380

Asn Phe Arg Gly Glu Lys Leu Pro Pro Val Ile Tyr Val Gly His Pro
385 390 395 400

Trp Asp Val Thr Arg Thr Val Glu Val Ser Ala Phe Val Pro Arg Tyr
405 410 415

Ser Leu Leu Lys Ala Ala Trp Pro Asp Ala Trp Leu Leu Leu Leu Lys
420 425 430

Tyr Asn Val Val Phe Ser Ile Gly Leu Ala Leu Ile Asn Ala Ile Pro
435 440 445

Cys Phe Gly Phe Asp Gly Ala His Ile Thr Ser Thr Val Ile His Ser
450 455 460

Phe Leu Val Gly Arg Val Asp Gln His Ala Lys Arg Asp Ile Ile Ser
465 470 475 480

Leu Ile Ile Thr Ser Val Gly Ser Leu Leu Phe Ala Leu Ala Leu Leu
485 490 495

Lys Val Ala Trp Leu Ser Phe Leu Arg Pro Leu Leu
500 505

<210> 5
<211> 4128
<212> DNA
<213> Drosophila melanogaster

<400> 5
gtgtgcctga ctgttttgta ggtgtaagga ggggcgtggc caaatagttt ttggtatacg 60
gatagaattt ggatgaaaaa taaaacgaaa tcaaaacatt tttcaaaagc gtggaagttt 120
tggccggctt gtgggcatgg caaaacgttt tttggctatc cgttaatcaa cataccgttg 180
cccgggacaa tacccaccaa gatcgttgta ccctacgaaa ctggatccgg atcgctgtca 240
tggcactctc ttaatacatc ctcgactaca ccgcaggaac cgcacccttc cggcgaaccc 300
tggccccccg aaccacaggt actcaatagc agtaccacgg accgcagccc gcctcccctt 360
ctgccctggg cgcagagcag ccccgccttt ttctacgtcc agcagattac tctgcgaacc 420
agtgttctcc cgtggacgga gggaatgcag cttatggatg cgtttcgtgc gccgctacac 480
gaagttttta aattgcttga aattgtgcgc aatcaccaga gcagcgaaaa caaacgtacc 540
ctggagcaca actgcctaca tgtagacaac gtaaagcgcg gaacacacgg gcagctggac 600
cagatctttc cggagtatgg ctgcctgctg ctctcgcccg ccaacctgtg gacgcagaac 660
tctcagaact ttactcggga cacaaacatc ctgaacacga tatttcagta ccataaccta 720
cagaaatcaa aagtttccgc ggcggaaatg ctgtttggat tacccatgca ggacactgga 780
ttcaagcgct atccattgcg cgctcggtcg cgtattatac agtatgcctt gacgttattc 840
ctcaagcaca acgatatgga gtatctggac actctaaagg aaaagctgct gcgacactat 900
cccccactcc cgttggctag tgcgtcggct gaagagccga cgaccataac ttacatcttt 960
tatccaggag agtacaggat gtgggagctg gtgccttaca cagtggcctt tatgttggtg 1020
tttgcttatg tgtacttctc tgttcgaaaa atcgatgtat ttcgttcccg ctttttgctg 1080
gccttatgta gcgtaatcac cacagccggg agcttggcca tgtcccttgg cttgtgtttc 1140
ttctttggcc tgacaatttc gctgcagtca aaggacattt tcccctacct tgtaatcctt 1200
gtgggattgg aaaatagctt ggtgatcaca aagagcgtag tctcaatgga cgagacattc 1260
gacgtgaaga tccgcgtggc gcaggctctt agcaaggagg gttggcatat atccaagact 1320
cttttgacgg agataacaat tttgacaatt ggtcttgcta ctttcgtgcc cgtcatccag 1380
gagttttgta tctttgccat agtcggcttg ctttccgatt ttatgctaca gatgctgctc 1440
ttctcaacaa tactggccat gaacattaag cggaccgagt atacggcgga ggccaagcac 1500

cttcctaaga tgttgctgag ctgcacccaa ggggctggtc gacaggattt ccgatttttc 1560
ggggccgccc cagcactgcc accgtttgtc cctggcacat ttcagcgttc tcagtcgcat 1620
ccaaaactgt gttttgctga tcccgcatct gttagcgatc gtacaagctt ggttaatgga 1680
cactcgtcgc cggagcaacg aatacccaaa cgcataaaga ttgtaaattt ctgggcgcgg 1740
actcgctttt ttcagcgtgc cttcatgatc tggatgattg tgtggatatg ctctatagtt 1800
tataattcgg gatatctgga gcagttgttt agcatgcaga gcaacggcac aatgacggca 1860
acccttgaac ttcaacggcg actacaggcg ggtcggggag cagtcagcag ttttttcgag 1920
ggatggcaag cggacgggca gcgtgccacg agtgcgccaa gcggaagcgg cttttctacg 1980
ccaataaaag ctcctctagc gatcgatata aacgaaacgg ccgaggaaat gatgagactt 2040
cgatatccca gcttcgacct aaactatttc ctttcaaact tccactggtc cacgattatg 2100
aaacagtaca acatctcact aagtgggcac tacgttaccc tgctaccgac cattcgcctt 2160
agtcatgcca tcgctccgga gctagccact ctgttgcgga atccgcagga gcagctgcaa 2220
caaaattttc aatggaaggc cctagccgct gcactcgatc cgctggactt taacgatgac 2280
gacgtgcgcc gtgagtctcc gatggtaatg gcagaggggt tgcctctggt tcccaagagc 2340
cccatggaaa tatttttcgc catcctcttg tgctgcatca gcatcttcgt gctttgctac 2400
acgatggtgg ttttctaccg ctgcatatgt accaggaact atgccgagtg gcgctccagt 2460
tggcacgaat ctgaggcacc gtacaagcag actgagcaaa tcctggaggg agttccaacg 2520
caaatcgccg gacacaaaca tcgcattgaa tgcctggtgt ctgacggcgc ctacataatc 2580
agctgctgcc ttaaaggcca aatccgagtg tgggatgcac gcagtggcga gcagctaacc 2640
agcatctccc gatccgatat tcagatctct cagcagcgga cggatgggca gacgctggta 2700
cgaaagctgg ccgtgtcacc ggtctggtgc cttgactact tcgataatct aatcgcagta 2760
ggctgcgcca acggccgcgt agaattgtgg gaatcccctg cgggattgct taagtgtgca 2820
taccaggaag acgcgaagag aaaccagggt ataacccaca tccacctgaa cggcgatcga 2880
gtgattgtgg cgcgtcttaa tggccgacta gatttttacc gcttagagac gtactacaag 2940
gggaagcaaa tcgactgggg ttttacctcg gcttacagga gaactcatgt tcgaactgga 3000
tccactggaa gcctgggatt aatgttgcag cagcagcgct gtcagcaaga agcatcccag 3060
aagaccacca aggaggaaat gaaaatcaca ttggagggtg taagactagc ccatcagcag 3120
ccaatcacat gcatgcaggt cgttaacgac atggttttca ctggcagcca ggatcacacc 3180
ctcaaggtgt attgcctcaa taagtcggat gttgagtata cgctccacgg tcactgtggg 3240
cctgtaacct gtctctttgt ggatcgctgg caacctggca caggggggtc tgggtcccag 3300
gacggcctgc tctgcgtatg ggatctgttc acgggagcct gcatgtataa tatacaagct 3360
cacgacggag ccgtcagctg cctggcctgt gcgcccagtt acgtaatctc gctaggcacg 3420
gacgagagga tttgcgtatg ggaacgattt cagggaaacc tgttgactac catcaacatc 3480
tcaaacgcat actcgagcct actgatgcta acaccgtcac tattggttac gagcaaaatg 3540
ggtaaggcct cattcttgat tgccaatata agagggacag taaataataa atttaattcc 3600
aacacaggat ctcttattgt gtgggatgtg cgcactgggc agccggctcg cgaggtcaaa 3660
ctggactttg caaacctgca gctctgtccc aaaataatga tgcttgcctg cgattcggta 3720
gtttgcgact acggaaatga gatccgcgtc gtccgctttc ctatcgtggc agacaagtgc 3780
cattaaagcg caaaatttta atttagcgtg gttcgctagc acctaggaat aagttgactt 3840
aaggctttaa aacgcctgga agtcattgac gcattcacta ttttatataa atatatacac 3900
tattagggtc cgcagcaact tacggtttta acacaagctg tacgtatctc atctctagaa 3960
ttttgtgtta gtttgtggac actaagtgta acagctacgc tccggtaggt taaggaacta 4020
aactaaatga atcagatata tacacatata ttttcgcgta attatataaa ctacatagtg 4080
tcttaaagcg cctcagccta atataaaatg actaaatgtt aaaataaa 4128

<210> 6
<211> 1237

<212> PRT
<213> Drosophila melanogaster

<400> 6

```
Met Lys Asn Lys Thr Lys Ser Lys His Phe Ser Lys Ala Trp Lys Phe
 1               5                  10                  15

Trp Pro Ala Cys Gly His Gly Lys Thr Phe Phe Gly Tyr Pro Leu Ile
            20                  25                  30

Asn Ile Pro Leu Pro Gly Thr Ile Pro Thr Lys Ile Val Val Pro Tyr
        35                  40                  45

Glu Thr Gly Ser Gly Ser Leu Ser Trp His Ser Leu Asn Thr Ser Ser
    50                  55                  60

Thr Thr Pro Gln Glu Pro His Pro Ser Gly Glu Pro Trp Pro Pro Glu
65                  70                  75                  80

Pro Gln Val Leu Asn Ser Ser Thr Thr Asp Arg Ser Pro Pro Pro Leu
                85                  90                  95

Leu Pro Trp Ala Gln Ser Ser Pro Ala Phe Phe Tyr Val Gln Gln Ile
            100                 105                 110

Thr Leu Arg Thr Ser Val Leu Pro Trp Thr Glu Gly Met Gln Leu Met
        115                 120                 125

Asp Ala Phe Arg Ala Pro Leu His Glu Val Phe Lys Leu Leu Glu Ile
    130                 135                 140

Val Arg Asn His Gln Ser Ser Glu Asn Lys Arg Thr Leu Glu His Asn
145                 150                 155                 160

Cys Leu His Val Asp Asn Val Lys Arg Gly Thr His Gly Gln Leu Asp
                165                 170                 175

Gln Ile Phe Pro Glu Tyr Gly Cys Leu Leu Leu Ser Pro Ala Asn Leu
            180                 185                 190

Trp Thr Gln Asn Ser Gln Asn Phe Thr Arg Asp Thr Asn Ile Leu Asn
        195                 200                 205

Thr Ile Phe Gln Tyr His Asn Leu Gln Lys Ser Lys Val Ser Ala Ala
    210                 215                 220

Glu Met Leu Phe Gly Leu Pro Met Gln Asp Thr Gly Phe Lys Arg Tyr
225                 230                 235                 240
```

Pro Leu Arg Ala Arg Ser Arg Ile Ile Gln Tyr Ala Leu Thr Leu Phe
245 250 255

Leu Lys His Asn Asp Met Glu Tyr Leu Asp Thr Leu Lys Glu Lys Leu
260 265 270

Leu Arg His Tyr Pro Pro Leu Pro Leu Ala Ser Ala Ser Ala Glu Glu
275 280 285

Pro Thr Thr Ile Thr Tyr Ile Phe Tyr Pro Gly Glu Tyr Arg Met Trp
290 295 300

Glu Leu Val Pro Tyr Thr Val Ala Phe Met Leu Val Phe Ala Tyr Val
305 310 315 320

Tyr Phe Ser Val Arg Lys Ile Asp Val Phe Arg Ser Arg Phe Leu Leu
325 330 335

Ala Leu Cys Ser Val Ile Thr Thr Ala Gly Ser Leu Ala Met Ser Leu
340 345 350

Gly Leu Cys Phe Phe Phe Gly Leu Thr Ile Ser Leu Gln Ser Lys Asp
355 360 365

Ile Phe Pro Tyr Leu Val Ile Leu Val Gly Leu Glu Asn Ser Leu Val
370 375 380

Ile Thr Lys Ser Val Val Ser Met Asp Glu Thr Phe Asp Val Lys Ile
385 390 395 400

Arg Val Ala Gln Ala Leu Ser Lys Glu Gly Trp His Ile Ser Lys Thr
405 410 415

Leu Leu Thr Glu Ile Thr Ile Leu Thr Ile Gly Leu Ala Thr Phe Val
420 425 430

Pro Val Ile Gln Glu Phe Cys Ile Phe Ala Ile Val Gly Leu Leu Ser
435 440 445

Asp Phe Met Leu Gln Met Leu Leu Phe Ser Thr Ile Leu Ala Met Asn
450 455 460

Ile Lys Arg Thr Glu Tyr Thr Ala Glu Ala Lys His Leu Pro Lys Met
465 470 475 480

Leu Leu Ser Cys Thr Gln Gly Ala Gly Arg Gln Asp Phe Arg Phe Phe
485 490 495

Gly Ala Ala Pro Ala Leu Pro Pro Phe Val Pro Gly Thr Phe Gln Arg
500 505 510

Ser Gln Ser His Pro Lys Leu Cys Phe Ala Asp Pro Ala Ser Val Ser
515 520 525

Asp Arg Thr Ser Leu Val Asn Gly His Ser Ser Pro Glu Gln Arg Ile
530 535 540

Pro Lys Arg Ile Lys Ile Val Asn Phe Trp Ala Arg Thr Arg Phe Phe
545 550 555 560

Gln Arg Ala Phe Met Ile Trp Met Ile Val Trp Ile Cys Ser Ile Val
565 570 575

Tyr Asn Ser Gly Tyr Leu Glu Gln Leu Phe Ser Met Gln Ser Asn Gly
580 585 590

Thr Met Thr Ala Thr Leu Glu Leu Gln Arg Arg Leu Gln Ala Gly Arg
595 600 605

Gly Ala Val Ser Ser Phe Phe Glu Gly Trp Gln Ala Asp Gly Gln Arg
610 615 620

Ala Thr Ser Ala Pro Ser Gly Ser Gly Phe Ser Thr Pro Ile Lys Ala
625 630 635 640

Pro Leu Ala Ile Asp Ile Asn Glu Thr Ala Glu Glu Met Met Arg Leu
645 650 655

Arg Tyr Pro Ser Phe Asp Leu Asn Tyr Phe Leu Ser Asn Phe His Trp
660 665 670

Ser Thr Ile Met Lys Gln Tyr Asn Ile Ser Leu Ser Gly His Tyr Val
675 680 685

Thr Leu Leu Pro Thr Ile Arg Leu Ser His Ala Ile Ala Pro Glu Leu
690 695 700

Ala Thr Leu Leu Arg Asn Pro Gln Glu Gln Leu Gln Gln Asn Phe Gln
705 710 715 720

Trp Lys Ala Leu Ala Ala Ala Leu Asp Pro Leu Asp Phe Asn Asp Asp
725 730 735

Asp Val Arg Arg Glu Ser Pro Met Val Met Ala Glu Gly Leu Pro Leu
740 745 750

Val Pro Lys Ser Pro Met Glu Ile Phe Phe Ala Ile Leu Leu Cys Cys
755 760 765

Ile Ser Ile Phe Val Leu Cys Tyr Thr Met Val Val Phe Tyr Arg Cys
770 775 780

Ile Cys Thr Arg Asn Tyr Ala Glu Trp Arg Ser Ser Trp His Glu Ser
785 790 795 800

Glu Ala Pro Tyr Lys Gln Thr Glu Gln Ile Leu Glu Gly Val Pro Thr
805 810 815

Gln Ile Ala Gly His Lys His Arg Ile Glu Cys Leu Val Ser Asp Gly
820 825 830

Ala Tyr Ile Ile Ser Cys Cys Leu Lys Gly Gln Ile Arg Val Trp Asp
835 840 845

Ala Arg Ser Gly Glu Gln Leu Thr Ser Ile Ser Arg Ser Asp Ile Gln
850 855 860

Ile Ser Gln Gln Arg Thr Asp Gly Gln Thr Leu Val Arg Lys Leu Ala
865 870 875 880

Val Ser Pro Val Trp Cys Leu Asp Tyr Phe Asp Asn Leu Ile Ala Val
885 890 895

Gly Cys Ala Asn Gly Arg Val Glu Leu Trp Glu Ser Pro Ala Gly Leu
900 905 910

Leu Lys Cys Ala Tyr Gln Glu Asp Ala Lys Arg Asn Gln Gly Ile Thr
915 920 925

His Ile His Leu Asn Gly Asp Arg Val Ile Val Ala Arg Leu Asn Gly
930 935 940

Arg Leu Asp Phe Tyr Arg Leu Glu Thr Tyr Tyr Lys Gly Lys Gln Ile
945 950 955 960

Asp Trp Gly Phe Thr Ser Ala Tyr Arg Arg Thr His Val Arg Thr Gly
965 970 975

Ser Thr Gly Ser Leu Gly Leu Met Leu Gln Gln Gln Arg Cys Gln Gln
980 985 990

Glu Ala Ser Gln Lys Thr Thr Lys Glu Glu Met Lys Ile Thr Leu Glu
995 1000 1005

Gly Val Arg Leu Ala His Gln Gln Pro Ile Thr Cys Met Gln Val Val
1010 1015 1020

Asn Asp Met Val Phe Thr Gly Ser Gln Asp His Thr Leu Lys Val Tyr
1025 1030 1035 1040

Cys Leu Asn Lys Ser Asp Val Glu Tyr Thr Leu His Gly His Cys Gly
1045 1050 1055

Pro Val Thr Cys Leu Phe Val Asp Arg Trp Gln Pro Gly Thr Gly Gly
1060 1065 1070

Ser Gly Ser Gln Asp Gly Leu Leu Cys Val Trp Asp Leu Phe Thr Gly
1075 1080 1085

Ala Cys Met Tyr Asn Ile Gln Ala His Asp Gly Ala Val Ser Cys Leu
1090 1095 1100

Ala Cys Ala Pro Ser Tyr Val Ile Ser Leu Gly Thr Asp Glu Arg Ile
1105 1110 1115 1120

Cys Val Trp Glu Arg Phe Gln Gly Asn Leu Leu Thr Thr Ile Asn Ile
1125 1130 1135

Ser Asn Ala Tyr Ser Ser Leu Leu Met Leu Thr Pro Ser Leu Leu Val
1140 1145 1150

Thr Ser Lys Met Gly Lys Ala Ser Phe Leu Ile Ala Asn Ile Arg Gly
1155 1160 1165

Thr Val Asn Asn Lys Phe Asn Ser Asn Thr Gly Ser Leu Ile Val Trp
1170 1175 1180

Asp Val Arg Thr Gly Gln Pro Ala Arg Glu Val Lys Leu Asp Phe Ala
1185 1190 1195 1200

Asn Leu Gln Leu Cys Pro Lys Ile Met Met Leu Ala Cys Asp Ser Val
1205 1210 1215

Val Cys Asp Tyr Gly Asn Glu Ile Arg Val Val Arg Phe Pro Ile Val
1220 1225 1230

Ala Asp Lys Cys His
1235

<210> 7

<211> 3160
<212> DNA
<213> Drosophila melanogaster

<400> 7
gcacgagcac agacgaagcc tgttgaactt ttcgtaaaaa ttattccttt cgcaaagtta 60
aatgaatgtg tttactttct tatttataat aagcgcaatt tgcagcctcg acgcctttaa 120
aacagccgtt gttccaaatg agttcatcgt tcacttccat tcaaaatact ttgccccggt 180
ccgagaatcc tacatcgcag caaaacttct tggttcaaac gtaacgaact ggagaattgt 240
tccccgtcta aatttggctt ggcaatatcc aagtgatttt gatatcttac gagtttgcga 300
cggttatgaa tcatcatcag agtttattat agaaaggctt cagactcacc catcagtaaa 360
ggcagtagtt ccccagcgaa gcgtacgaag gatcctaaac tatgacgcct atagcaacct 420
aacgtatatt caccgccatc cccaaggagt gctaaggaac agaaacccaa acaacgatcg 480
ccaccgacaa ttgtgctccg tactccacgc caacatcctt tggaagctgg gtatcacagg 540
caagggagtt aaagtggcca ttttcgacac tggcctaacc aaaaaccatc cacactttcg 600
aaatgtaaag gaacgaacaa actggacgaa tgaaaagtca cttgacgaca gagtcagtca 660
tggcaccttc gtcgccgggg taatcgcttc ttccagggaa tgcctaggct tcgctcccga 720
cgccgatctt tacatattta aagtttttac gaactcccaa gtttcttaca cttcctggtt 780
cctggatgca ttcaactacg cgatatatag gaaaataaac attctcaacc ttagcattgg 840
gggtcccgac tttatggact cgccgttcgt tgaaaaggtg ttggaactgt cggctaataa 900
tgtcataatg atatcggcag caggaaatga tggtcccttg tacggcacgc taaacaatcc 960
tggcgatcag agcgatgtag ttggcgttgg tggcattcag tttgatgata aaatcgccaa 1020
gtttagttcg agaggaatga caacgtggga acttcccttа ggctacggac gtatgggact 1080
cgatattgtc acgtacggaa gtcaagtgga aggcagtgat gtgcgcaagg ggtgcagacg 1140
actctctgga acatccgtgt cctctccagt tgttgcaggg gctgctgcac tgcttataag 1200
cggtgcattt cagaaaatcg actacataaa cccagcatct cttaagcagg tactcattga 1260
aggtgccgag aaactgccgc attataacat gtttgagcag ggagctggaa aactgaattt 1320
gctgaagagt atgcagctat tgctgtcata caaaccaaag ataaccctta ttccggcata 1380
ccttgacttc acccaaaact atatgtggcc ttatagctcc caacctctgt actatggaag 1440
ctccgtcgct attgcaaacg ttaccatact caatggtatc tctgtcacaa gtcatatagt 1500
tggcatccct aaatggattc ccgatttcga aaaccaaggt cagtttcttc aagtatctgc 1560
acaagtttcg cctatcgttt ggccgtggac cggttggatg tcagttttta ttgctgtaaa 1620
aaaggaagga gaaaactttg aaggtgtttg taaaggaagt atcaccctag ttttggaaag 1680
ctttaaacag accaccaacg aaactcatgt tacagaagtc gactttcctt taacaataaa 1740
ggttactcca aaaccgccaa gaaacaagag gattttatgg gatcagtacc acagcctaag 1800
gtatccaccg cgctatattc cacgagatga tctcaaagtt aaactagatc ctctggactg 1860
gagggcagac catatacaca caaactttag ggacatgtat acacatttac gaaatgttgg 1920
ctactacatt gatgttttgc gagaaccctt cacctgcttc aatgcctcgg attatggcgc 1980
gttattgatt gttgaccctg agagagggtt tggcgacgag gaaataaacg ctttacagga 2040
aaacgtgtat aaaagaggct tgaatgtcgt cgtattcgga gactggtata acaccactgt 2100
gatgaaaaaa attaaattct ttgacgagaa cacccgacaa tggtggacac ccgacactgg 2160
tggcgcaaat attccagcct tgaatgattt attgaagcca tttggaattg cttttggcga 2220
ttttgtcggt gagggacatt tcaaactggg cgaccattca atgtactatg ctagtggagc 2280
cacaattgtt aagtttccaa tgaatccagg agatattata gtgggcacaa aactgaatga 2340
ccaaggactt tcgattatta attctaaaac acccagcaag gtagcaaaac tagatgtacc 2400
tatttttggt atgttccaaa ccaaggcgaa cagtattcaa agcaacgagg aaatcgtggt 2460
caatgcggaa agcaatttgg cagaggctat acccacagat tactccacat ttaagaaccg 2520
ggttttgcta ctgcgaacga agcaacgaag tatcagtttt gcgaaaagca ataatcatga 2580

```
aactaagaat gaaggacgta ttgccgtata tggggactcc aactgcctcg actccacgca 2640
tctggagaag gcttgctact ggctgctaat aacgttttta gattttgcaa taaactcgca 2700
caaatcaagt ttattgcaga atctaaatcg tataactgaa tttcacaaat tagagagagc 2760
accattaccc cttaggatat cgcaaagtat tataaaatct cgttcacagg acaataattg 2820
tgaacaattt aagtggcttg caccgacgaa gcaaaataac gccgaggaaa ggaaatcttc 2880
tataatagac gtaaccatac tggaaaatga agaacacgag ataaatttaa tcaaaaattt 2940
attgggtgag gagatcgcaa aactagggca aaacaatgat tatttaacag gaatgcaatc 3000
cgcggatagt ctaatgactc caatatattc taattatnat aagcctaant gtatcatgta 3060
tttgtttatt ttaaagcgtc attcatttgt aataatactt cactttaaaa cgtaaaaaaa 3120
aaaaaactga ggggccgtac cattcgctaa ggagcgatct                      3160
```

<210> 8
<211> 993
<212> PRT
<213> Drosophila melanogaster

<400> 8

```
Met Asn Val Phe Thr Phe Leu Phe Ile Ile Ser Ala Ile Cys Ser Leu
 1               5                  10                  15

Asp Ala Phe Lys Thr Ala Val Val Pro Asn Glu Phe Ile Val His Phe
            20                  25                  30

His Ser Lys Tyr Phe Ala Pro Val Arg Glu Ser Tyr Ile Ala Ala Lys
        35                  40                  45

Leu Leu Gly Ser Asn Val Thr Asn Trp Arg Ile Val Pro Arg Leu Asn
    50                  55                  60

Leu Ala Trp Gln Tyr Pro Ser Asp Phe Asp Ile Leu Arg Val Cys Asp
65                  70                  75                  80

Gly Tyr Glu Ser Ser Ser Glu Phe Ile Ile Glu Arg Leu Gln Thr His
                85                  90                  95

Pro Ser Val Lys Ala Val Val Pro Gln Arg Ser Val Arg Arg Ile Leu
            100                 105                 110

Asn Tyr Asp Ala Tyr Ser Asn Leu Thr Tyr Ile His Arg His Pro Gln
        115                 120                 125

Gly Val Leu Arg Asn Arg Asn Pro Asn Asn Asp Arg His Arg Gln Leu
    130                 135                 140

Cys Ser Val Leu His Ala Asn Ile Leu Trp Lys Leu Gly Ile Thr Gly
145                 150                 155                 160
```

Lys Gly Val Lys Val Ala Ile Phe Asp Thr Gly Leu Thr Lys Asn His
165 170 175

Pro His Phe Arg Asn Val Lys Glu Arg Thr Asn Trp Thr Asn Glu Lys
180 185 190

Ser Leu Asp Asp Arg Val Ser His Gly Thr Phe Val Ala Gly Val Ile
195 200 205

Ala Ser Ser Arg Glu Cys Leu Gly Phe Ala Pro Asp Ala Asp Leu Tyr
210 215 220

Ile Phe Lys Val Phe Thr Asn Ser Gln Val Ser Tyr Thr Ser Trp Phe
225 230 235 240

Leu Asp Ala Phe Asn Tyr Ala Ile Tyr Arg Lys Ile Asn Ile Leu Asn
245 250 255

Leu Ser Ile Gly Gly Pro Asp Phe Met Asp Ser Pro Phe Val Glu Lys
260 265 270

Val Leu Glu Leu Ser Ala Asn Asn Val Ile Met Ile Ser Ala Ala Gly
275 280 285

Asn Asp Gly Pro Leu Tyr Gly Thr Leu Asn Asn Pro Gly Asp Gln Ser
290 295 300

Asp Val Val Gly Val Gly Gly Ile Gln Phe Asp Asp Lys Ile Ala Lys
305 310 315 320

Phe Ser Ser Arg Gly Met Thr Thr Trp Glu Leu Pro Leu Gly Tyr Gly
325 330 335

Arg Met Gly Leu Asp Ile Val Thr Tyr Gly Ser Gln Val Glu Gly Ser
340 345 350

Asp Val Arg Lys Gly Cys Arg Arg Leu Ser Gly Thr Ser Val Ser Ser
355 360 365

Pro Val Val Ala Gly Ala Ala Ala Leu Leu Ile Ser Gly Ala Phe Gln
370 375 380

Lys Ile Asp Tyr Ile Asn Pro Ala Ser Leu Lys Gln Val Leu Ile Glu
385 390 395 400

Gly Ala Glu Lys Leu Pro His Tyr Asn Met Phe Glu Gln Gly Ala Gly
405 410 415

Lys Leu Asn Leu Leu Lys Ser Met Gln Leu Leu Leu Ser Tyr Lys Pro
420 425 430

Lys Ile Thr Leu Ile Pro Ala Tyr Leu Asp Phe Thr Gln Asn Tyr Met
435 440 445

Trp Pro Tyr Ser Ser Gln Pro Leu Tyr Tyr Gly Ser Ser Val Ala Ile
450 455 460

Ala Asn Val Thr Ile Leu Asn Gly Ile Ser Val Thr Ser His Ile Val
465 470 475 480

Gly Ile Pro Lys Trp Ile Pro Asp Phe Glu Asn Gln Gly Gln Phe Leu
485 490 495

Gln Val Ser Ala Gln Val Ser Pro Ile Val Trp Pro Trp Thr Gly Trp
500 505 510

Met Ser Val Phe Ile Ala Val Lys Lys Glu Gly Glu Asn Phe Glu Gly
515 520 525

Val Cys Lys Gly Ser Ile Thr Leu Val Leu Glu Ser Phe Lys Gln Thr
530 535 540

Thr Asn Glu Thr His Val Thr Glu Val Asp Phe Pro Leu Thr Ile Lys
545 550 555 560

Val Thr Pro Lys Pro Pro Arg Asn Lys Arg Ile Leu Trp Asp Gln Tyr
565 570 575

His Ser Leu Arg Tyr Pro Pro Arg Tyr Ile Pro Arg Asp Asp Leu Lys
580 585 590

Val Lys Leu Asp Pro Leu Asp Trp Arg Ala Asp His Ile His Thr Asn
595 600 605

Phe Arg Asp Met Tyr Thr His Leu Arg Asn Val Gly Tyr Tyr Ile Asp
610 615 620

Val Leu Arg Glu Pro Phe Thr Cys Phe Asn Ala Ser Asp Tyr Gly Ala
625 630 635 640

Leu Leu Ile Val Asp Pro Glu Arg Gly Phe Gly Asp Glu Glu Ile Asn
645 650 655

Ala Leu Gln Glu Asn Val Tyr Lys Arg Gly Leu Asn Val Val Val Phe
660 665 670

Gly Asp Trp Tyr Asn Thr Thr Val Met Lys Lys Ile Lys Phe Phe Asp
675 680 685

Glu Asn Thr Arg Gln Trp Trp Thr Pro Asp Thr Gly Gly Ala Asn Ile
690 695 700

Pro Ala Leu Asn Asp Leu Leu Lys Pro Phe Gly Ile Ala Phe Gly Asp
705 710 715 720

Phe Val Gly Glu Gly His Phe Lys Leu Gly Asp His Ser Met Tyr Tyr
725 730 735

Ala Ser Gly Ala Thr Ile Val Lys Phe Pro Met Asn Pro Gly Asp Ile
740 745 750

Ile Val Gly Thr Lys Leu Asn Asp Gln Gly Leu Ser Ile Ile Asn Ser
755 760 765

Lys Thr Pro Ser Lys Val Ala Lys Leu Asp Val Pro Ile Phe Gly Met
770 775 780

Phe Gln Thr Lys Ala Asn Ser Ile Gln Ser Asn Glu Glu Ile Val Val
785 790 795 800

Asn Ala Glu Ser Asn Leu Ala Glu Ala Ile Pro Thr Asp Tyr Ser Thr
805 810 815

Phe Lys Asn Arg Val Leu Leu Leu Arg Thr Lys Gln Arg Ser Ile Ser
820 825 830

Phe Ala Lys Ser Asn Asn His Glu Thr Lys Asn Glu Gly Arg Ile Ala
835 840 845

Val Tyr Gly Asp Ser Asn Cys Leu Asp Ser Thr His Leu Glu Lys Ala
850 855 860

Cys Tyr Trp Leu Leu Ile Thr Phe Leu Asp Phe Ala Ile Asn Ser His
865 870 875 880

Lys Ser Ser Leu Leu Gln Asn Leu Asn Arg Ile Thr Glu Phe His Lys
885 890 895

Leu Glu Arg Ala Pro Leu Pro Leu Arg Ile Ser Gln Ser Ile Ile Lys
900 905 910

Ser Arg Ser Gln Asp Asn Asn Cys Glu Gln Phe Lys Trp Leu Ala Pro
915 920 925

Thr Lys Gln Asn Asn Ala Glu Glu Arg Lys Ser Ser Ile Ile Asp Val
930 935 940

Thr Ile Leu Glu Asn Glu Glu His Glu Ile Asn Leu Ile Lys Asn Leu
945 950 955 960

Leu Gly Glu Glu Ile Ala Lys Leu Gly Gln Asn Asn Asp Tyr Leu Thr
965 970 975

Gly Met Gln Ser Ala Asp Ser Leu Met Thr Pro Ile Tyr Ser Asn Tyr
980 985 990

Xaa